Multidimensional characterization of quality of experience of stereoscopic 3D TV

Wei Chen

l'université nantes angers le mans
PÔLE DE RECHERCHE ET D'ENSEIGNEMENT SUPÉRIEUR

UNIVERSITÉ DE NANTES

Thèse de Doctorat

Wei CHEN

Mémoire présenté en vue de l'obtention
du grade de Docteur de l'Université de Nantes
Sous le label de l'Université Nantes Angers Le Mans

Discipline : *Informatique*
Spécialité : *Automatique et Informatique Appliquée*
Laboratoire : *IRCCyN*

Soutenue le 23 octobre 2012

École doctorale : 503 (STIM)
Thèse n° : ED503-179

Caractérisation multidimensionnelle de la qualité d'expérience en télévision de la TV3D stéréoscopique

Multidimensional characterization of quality of experience of stereoscopic 3D TV

JURY

Rapporteurs :

Lina Karam, Professeur, Arizona State University

Christophe Charrier, Maître de Conférences HDR, GREYC, Université de Caen Basse Normandie

Examinateurs :

Marcus Barkowsky, Maître de Conférences, IRCCyN, Université de Nantes

Luce Morin, Professeur des Universités, IETR, INSA de Rennes

Touradj Ebrahimi, Professeur, Ecole Polytechnique Fédérale de Lausanne

Invité : **Jérôme Fournier**, ingénieur expert, lab'Orange, France Télécom

Directeur de Thèse : **Patrick le Callet**, Professeur des Universités, IRCCyN, Université de Nantes

L'UNAM — l'université nantes angers le mans
PÔLE DE RECHERCHE ET D'ENSEIGNEMENT SUPÉRIEUR

UNIVERSITÉ DE NANTES

Thèse de Doctorat

Wei CHEN

Caractérisation multidimensionnelle de la qualité d'expérience de la TV3D stéréoscopique

Multidimensional characterization of quality of experience of stereoscopic 3D TV

Résumé

La TV 3D stéréoscopique (S-3DTV) est supposée améliorer la sensation de profondeur des observateurs mais possiblement en affectant d'autres facteurs de l'expérience utilisateur. L'évaluation subjective (avec observateurs) est la méthode la plus directe pour qualifier la qualité d'expérience (QoE). Cependant, les méthodes conventionnelles ne sont pas adaptées à l'évaluation de la QoE dans le cas de la S-3DTV. Cette thèse a pour but de, premièrement proposer de nouvelles méthodologies pour évaluer la QoE dans pareil contexte ; deuxièmement investiguer les impacts de choix technologiques de la diffusion S-3DTV sur la QoE ; troisièmement proposer des recommandations pour optimiser la QoE. Sur les aspects méthodologiques, l'idée clé repose sur une approche multidimensionnelle de la QoE via la définition de plusieurs indicateurs. La fatigue visuelle fait l'objet d'une étude expérimentale particulière en utilisant des questionnaires, tests de vision et analyse de signaux EEG dans des conditions de visualisation optimisés. D'autres indicateurs ont été mesurés pour investiguer quantitativement l'impact de l'acquisition, la représentation, la compression et la transmission du contenu S-3DTV sur la QoE. De plus, les règles améliorées de captation stéréoscopiques, de budget de profondeur «confortable», de débit de diffusion ont été élaborées et validées au travers des études expérimentales.

Mots clés

TV 3D, qualité d'expérience, fatigue visuelle, confort visuel, qualité d'image, perception visuelle humaine, diffusion 3D

Abstract:

Stereoscopic-3DTV (S-3DTV) should provide enhanced depth perception to viewer while it might affect other factors of user experience. Subjective assessment is the most direct way to assess quality of experience (QoE). However, conventional assessment methods are not sufficient to evaluate the QoE of S-3DTV. This thesis aims first to propose new methodologies to evaluate S-3DTV QoE; second, investigate different technical issues related to QoE along the 3DTV broadcasting chain; third, propose recommendations to optimize the S-3DTV QoE. For methodological aspects, the key idea relies on using multidimensional QoE indicators. Visual fatigue, as a particular dimension of QoE, is addressed separately under optimized viewing conditions using questionnaire, vision test and EEG signals. For other QoE indicators, we design subjective QoE experiments to investigate the impact of content acquisition, 3D representation format, compression and transmission on QoE of S-3DTV. The experiment results quantitatively reveal how perceived binocular depth, compression distortion, the cooperation between 3D representation formats and line interleaved display, and view asymmetries affect multidimensional QoE of S-3DTV. Additionally, we elaborate and validate improved stereoscopic shooting rules, depth budget for visual comfort, appropriate frame compatible format for line interleaved display, bitrate to broadcast S-3DTV, threshold for view asymmetries to avoid visual discomfort.

Key words

3DTV, quality of experience, visual fatigue, visual comfort, image quality, human visual perception, 3D broadcasting

L'UNIVERSITÉ NANTES ANGERS LE MANS

Acknowledgements

The work presented in this thesis could not have been possible without the support of mange people.

Many thanks to my supervisors: Patrick Le Callet, Marcus Barkowsky and Jérôme Fournier for their timely advice, consultations, encouragement, and critiques throughout the development of this work. Jérôme has been an invaluable source of support and guidance all along my work on the thesis.

I am very grateful to France telecom colleagues: Bernard Letertre and Jean-Chareles Gicquel, for their discussion and help in designing and guiding subjective quality assessment experiments.

Last but not least, I would like to thank my family: my wife and my parents. They have always been a source of motivation. Without them, I could have never accomplished what I have done today.

Wei Chen

Abstract

Stereoscopic-3DTV (S-3DTV) should provide enhanced depth perception to viewer while it might affect other factors of user experience. Subjective assessment is the most direct way to assess quality of experience (QoE). However, conventional assessment methods are not sufficient to evaluate the QoE of S-3DTV.

This thesis aims first to propose new methodologies to evaluate S-3DTV QoE; second, investigate different technical issues related to QoE along the 3DTV broadcasting chain; third, propose recommendations to optimize the S-3DTV QoE.

For methodological aspects, the key idea relies on using multidimensional QoE indicators. Visual fatigue, as a particular dimension of QoE, is addressed separately under optimized viewing conditions using questionnaire, vision test and EEG signals. For other QoE indicators, we design subjective QoE experiments to investigate the impact of content acquisition, 3D representation format, compression and transmission on the QoE of S-3DTV. The experiment results quantitatively reveal how perceived binocular depth, compression distortion, the cooperation between 3D representation formats and line interleaved display, and view asymmetries affect multidimensional QoE of S-3DTV. Additionally, we elaborate and validate improved stereoscopic shooting rules, depth budget for visual comfort, appropriate frame compatible format for line interleaved display, bitrate to broadcast S-3DTV, threshold for view asymmetries to avoid visual discomfort.

Keywords: 3DTV, quality of experience, visual fatigue, visual comfort, image quality, human visual perception, 3D broadcasting

Résumé

La TV 3D stéréoscopique (S-3DTV) est supposée améliorer la sensation de profondeur des observateurs mais possiblement en affectant d'autres facteurs de l'expérience utilisateur. L'évaluation subjective (avec observateurs) est la méthode la plus directe pour qualifier la qualité d'expérience (QoE). Cependant, les méthodes conventionnelles ne sont pas adaptées à l'évaluation de la QoE dans le cas de la S-3DTV.

Cette thèse a pour but de, premièrement proposer de nouvelles méthodologies pour évaluer la QoE dans pareil contexte ; deuxièmement investiguer les impacts de choix technologiques de la diffusion S-3DTV sur la QoE ; troisièmement proposer des recommandations pour optimiser la QoE.

Sur les aspects méthodologiques, l'idée clé repose sur une approche multidimensionnelle de la QoE via la définition de plusieurs indicateurs. La fatigue visuelle fait l'objet d'une étude expérimentale particulière en utilisant des questionnaires, tests de vision et analyse de signaux EEG dans des conditions de visualisation optimisés. D'autres indicateurs ont été mesurés pour investiguer quantitativement l'impact de l'acquisition, la représentation, la compression et la transmission du contenu S-3DTV sur la QoE. De plus, les règles améliorées de captation stéréoscopiques, de budget de profondeur «confortable», de débit de diffusion ont été élaborées et validées au travers des études expérimentales.

Mots clés: TV 3D, qualité d'expérience, fatigue visuelle, confort visuel, qualité d'image, perception visuelle humaine, diffusion 3D

Contents

List of Figures

List of Tables

General Introduction

I. The history of stereoscopic images

The stereoscopic images history can be traced back to the first description of stereoscopic vision by Euclid (280 B.C). He described that the depth perception is obtained when each eye simultaneously perceives two slightly different images of the same object. In 1838, Sir Charles Wheatstone (Wheatstone, 1838) invented the first stereoscopic viewing device – the stereoscope as shown in Figure I- 1. The basic idea of this device was to separate the left and right viewing channels by additional instruments, e.g., mirrors, and to present different images individually to left and right eyes. This is also the basic principle and ancestor of modern stereoscopic device.

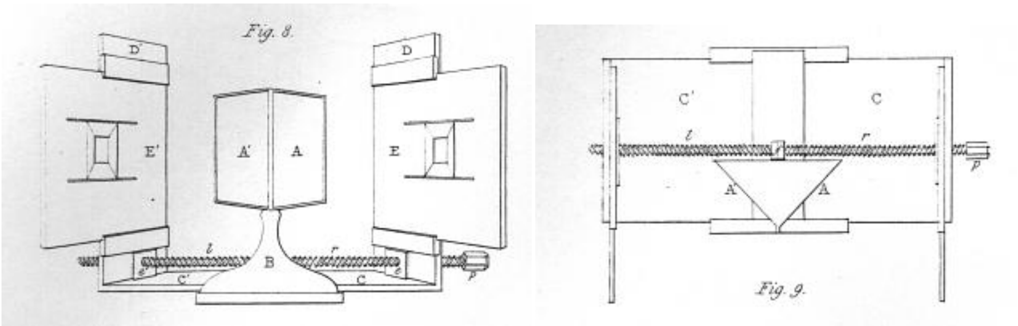


Figure I- 1 : The lenticular stereoscope (Wheatstone, 1838)

Between the 1840s and 1920s, stereoscopic images served as an important method of entertainment, education and virtual travel – predecessors to contemporary forms of media such as television and movies (Spiro, n.d.).

With the rapid development of modern movie and television technology, the first 3D test movie in anaglyph was produced by Edwin S. Porter and William E. Waddell in 1915. In 1922, the first public 3D movie in anaglyph “The power of love” was premiered (Zone, 2007) at the Ambassador hotel theatre in Los Angeles, American. In 1928, stereoscopic television was demonstrated for the first time by John Logie Baird. Later on, Edwin H. Land invented a polarizing sheet called Polaroid in 1932 and thereafter the polarization view separation technique started to be used to present stereoscopic movies as it can provide better quality than anaglyph technique.

In the 1950s, when TV became popular, many 3D movies were produced. The 1952 to 1955 period is called the first “golden era” for stereoscopic movies industry starting from the first colour stereoscopic feature, “Bwana Devil” presented to publics by Polaroid technique. A string of successful 3D movies was produced in this era. For example, the very first cartoon in 3D “Melody” by Walt Disney and the very first 3D movie with stereophonic sound “House of Wax” by Warner Bros were both produced and presented to the public during this era. However, the first 3D “golden era” declined from 1953 due to many reasons but mainly the immaturity of the production and display technology. For example, 3D required to project two synchronized prints simultaneously on the screen. If one print is broken, it is hard to maintain synchronization after repair. Moreover, for 3D based on Polaroid technique, the silver screen for reflecting the polarized wave was directional and caused side-line seating to be unusable with both 3D and regular films. By the mid to late of 1950's, 3D movies were out of favour and widescreen features were the dominant film format for moviegoers ("3D Moive Gaze", n.d.).

The revival of 3D started in the early 1960s with the invention of the Space-Vision 3D technique. In this technique, stereoscopic films were printed with two images, one above the other, in a single academy ratio frame, on a single strip. Thus, only one projector fitted with a special lens was needed (Mead, 2010). This so-called "over and under" technique re-attracted the producer and cinema owner back to 3D because it only required one projector and a broken print can still provide perfect synchronization after repair. In the 1960 to 1984, the main stream of stereoscopic 3D images in the cinema was still based on the anaglyph technology, which delivers the left and right images by separated colour channels. In 1985 to 2003, 3D display technologies based on the polarized glasses and active shutter glasses, which can provide better quality than anaglyph technology, was becoming more and more popular, e.g., the IMAX-3D cinema which has the capacity to record and display images of far greater size and resolution than conventional film systems.

By entering the 21st century, thanks to the rapid development of modern semiconductors and digital electronics technologies, stereoscopic 3D images resurged. In the cinema domain, combining with the computer rendering and editing technologies, more and more stereoscopic 3D movies were produced. One of the remarkable sign is that in 2009, the highest-grossing film of all time, AVATAR was presented mainly in stereoscopic 3D to the public. In the television domain, after the success and standardization of High Definition television (HDTV), the stereoscopic 3D television is widely discussed as the possible successor. The market research firm "Park association" estimated that 80% of TVs sold in 2014 will be capable of playing 3D content (Macchiarella, 2010).

II. Aim of this thesis

As presented in the previous section, stereoscopic 3D television (S-3DTV) might be the possible successor of HDTV. Compared with conventional 2DTV, the interest of S-3DTV is that it can provide enhanced depth sensation to viewers. However, it is still not a perfect representation of the real world and somehow it is only an illusion. Thus, new issues such as visual discomfort or stereoscopic distortion might be induced due to perceptual and/or technical problems.

Quality of experience (QoE) is a measure of customer's experience. "Picture Quality" is often used to represent the QoE for 2DTV. Subjective quality assessment is the conventional way to evaluate the "Picture Quality" of 2DTV system. However, first, "Picture Quality" is not sufficient to represent QoE of S-3DTV because it cannot directly highlight the advantages such as enhanced depth perception and the problems such as visual discomfort of S-3DTV. Second, conventional subjective quality assessment methods do not consider the new characteristics of S-3DTV, e.g., there is a lack of specification of the viewing environment for S-3DTV. Thus, developing new subjective QoE assessment methodologies dedicated to S-3DTV is mandatory. It will help to characterize the QoE of S-3DTV, ease the specification of end-to-end applications and optimize the design of different techniques for S-3DTV broadcasting.

The aim of this thesis covers three parts:

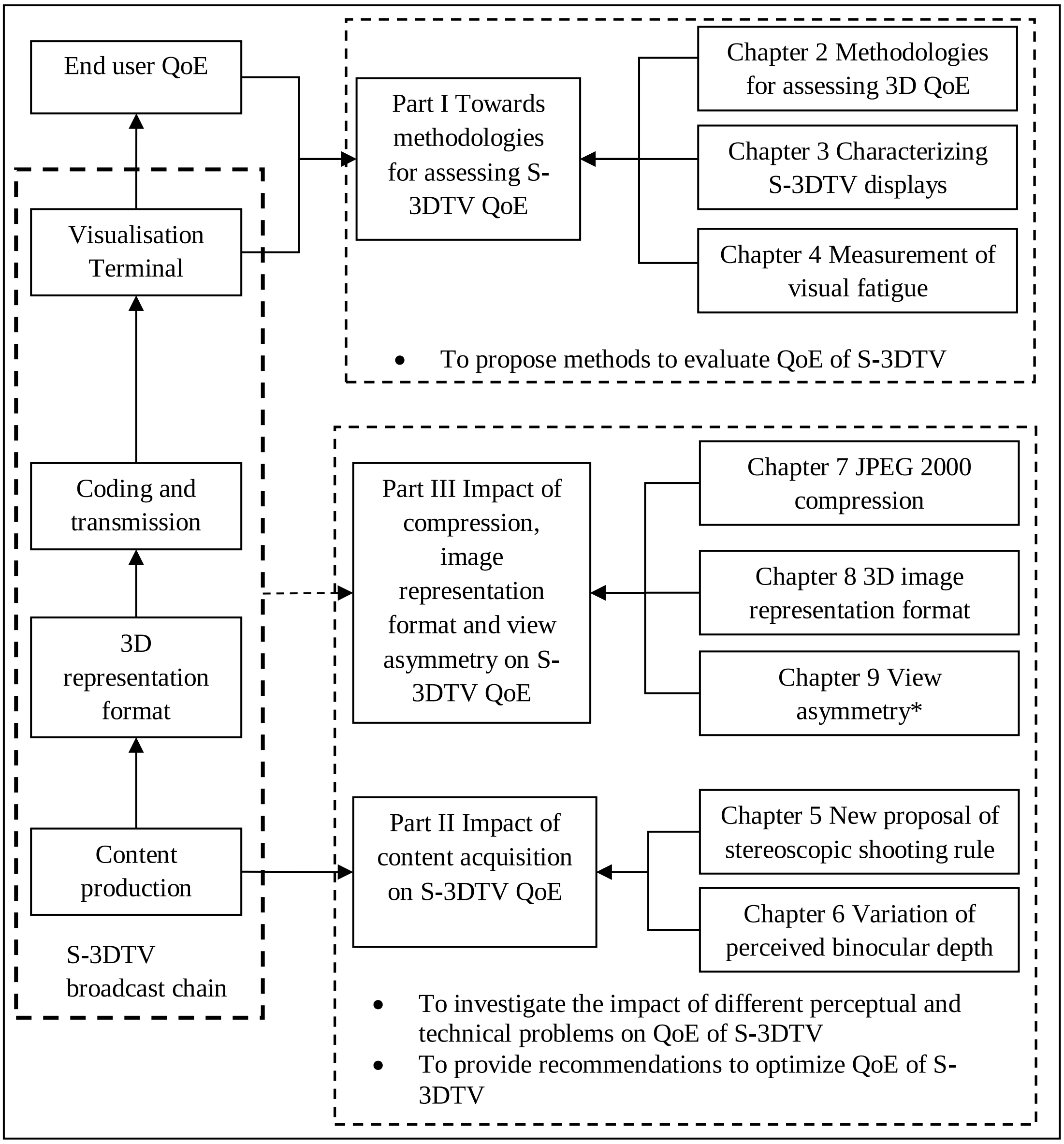
- To propose new methods to evaluate the QoE of S-3DTV
- To use the proposed methods to investigate the impact of different perceptual and technical problems (along the 3DTV broadcast chain) on the QoE of S-3DTV
- To provide recommendations related to perceptual and technical problems in order to optimize the QoE of S-3DTV

III. Overview of this thesis

Chapter 1 introduces the QoE challenges of S-3DTV as the background of this thesis. It presents the foundation of depth perception and the principle of stereoscopic imaging system. The fundamental advantages such as enhanced depth perception and problems such as visual discomfort and visual fatigue of S-3DTV on QoE are revealed. Moreover, the QoE issues related to different individual parts of the S-3DTV broadcasting chain (Content production, 3D representation format, coding and transmission and visualization terminal) are presented and discussed.

After this introduction, the contributions of the thesis are divided into three parts as illustrated in Figure I- 2. Each part is corresponding to different individual parts of the S-3DTV broadcasting chain.

Part I consists of three chapters (Chapter 2, Chapter 3, and Chapter 4). It presents the contributions of this thesis towards methodologies for assessing 3D QoE. In Chapter 2, first, we review the ITU recommendations and explorative studies related to subjective QoE assessment for S-3DTV. Second, towards a comprehensive adaption of subjective QoE assessment for S-3DTV, we propose to use multi-dimensional QoE indicators and to consider new factors affecting the QoE of S-3DTV in subjective assessment. Subjective QoE assessment with Multi-dimensional QoE indicators will serve as the main method for QoE assessment in this thesis. As display performance in subjective assessment is a critical issue affecting the QoE of S-3DTV, in Chapter 3, we propose new methods to characterize the luminance rendering and depth rendering of S-3DTV. Furthermore, Chapter 4 presents a study of measuring visual fatigue in optimal viewing condition. Three methods including vision test, questionnaire and EEG signal measurement are used in this study to measure visual fatigue.



* View asymmetry is a global problem related to every part of S-3DTV.

Figure I- 2 : Overview of contributions of this thesis

Part II including two chapters (Chapter 5 and Chapter 6) presents the contributions of this thesis towards understanding the impact of content acquisition on S-3DTV QoE. In Chapter 5, we propose stereoscopic shooting rules to optimize the content acquisition of S-3DTV considering stereoscopic distortion and the comfortable viewing zone in the final perception. Synthetic contents in different conditions corresponding to our improved shooting rules are generated. A subjective assessment with three QoE indicators is used to verify our improved shooting rules. In Chapter 6, both synthetic contents and natural contents in different levels of perceived binocular depth are generated controlling precisely shooting parameters. A subjective QoE assessment using six QoE indicators is carried out to evaluate the impact of variation of perceived binocular depth on the QoE of S-3DTV. Finally, a limit for perceived depth range is recommended.

Part III including three chapters (Chapter 7, Chapter 8 and Chapter 9) presents the contributions of this thesis to evaluate the impact of other important technical problems including compression, image representation format and view asymmetry on the QoE of S-3DTV. Chapter 7 focuses on the impact of JPEG 2000 compression on stereoscopic still images. Five QoE indicators are used in the subjective QoE assessment. Chapter 8 describes two experiments which aim at investigating the impact of 3D representation formats on the QoE of line interleaved S-3DTV. The first experiment focuses on understanding the resolution reduction effect of different frame-compatible formats on S-3DTV. The second experiment is designed to compare the QoE of different frame-compatible formats under different compression bitrates. Chapter 9 aims to evaluate the impact of view asymmetry on the QoE of S-3DTV. Perceptual thresholds for different types of view asymmetries are measured and recommended.

Chapter 1 QoE challenges for S-3DTV

Contents

1.1 Introduction

Quality of Experience (QoE) is a measure of customer's experiences. For S-3DTV, it is the measure of a viewer's experiences with stereoscopic images on S-3DTV. Compared with 2DTV, S-3DTV is able to provide additional depth information, i.e., the binocular disparity. This may enhance the depth perception and improve the QoE. Meanwhile, S-3DTV is still not a perfect presentation of a natural scene. Viewing stereoscopic images on S-3DTV may not be exactly the same as viewing a natural scene. These discrepancies may induce QoE issues and even result in visual discomfort and visual fatigue. Moreover, technical issues from the modern S-3DTV broadcast chain also have potential influence on the QoE of S-3DTV. In this chapter, we aim to present the QoE challenges for S-3DTV.

This chapter is organized as follows:

Section 1.2 presents the foundation of human depth perception. Different depth cues and their utilities at different depth ranges are introduced. Moreover, we present a focused discussion on binocular disparity which is the most important added value of S-3DTV. Section 1.3 presents the principle of stereoscopic imaging system which consists of image acquisition and image visualization. The discrepancies between viewing stereoscopic images and viewing real scenes are revealed. These discrepancies may result in visual discomfort and visual fatigue. Thus, Section 1.4

presents the potential impact of S-3DTV on visual discomfort and visual fatigue. Furthermore, different techniques in modern S-3DTV broadcast chains may also have potential influence on the QoE. Section 1.5 presents the characteristics of different techniques in the production, different formats of 3D representation, different coding and network transmission scenarios, and different visualization terminals. Moreover, their QoE issues are discussed. Section 1.6 summarizes the QoE challenges for S-3DTV.

1.2 Foundation of depth perception

Human depth perception, also called perception of layout, is the ability to see and understand the three-dimensional world. It is one of the major functions of our visual system. Since our eyes only have two-dimensional retinal images and no special third component for depth perception, it is an interpretation of physiological cues that leads to useful perception. Depth perception is the combination of the retinal images from our two eyes to extract the best and most convincing information about the three dimensions of our world. Strictly speaking, observers do not see depth but objects in depth, and they do not see space but objects in space.

Section 1.2.1 introduces different depth cues. Compared with 2DTV, S-3DTV adds stereoscopic information, i.e., the binocular disparity. Section 1.2.2 gives a focused discussion on the binocular disparity and how our visual system processes it to generate the 3D sensation. Section 1.2.3 discusses the sensitivity of different depth cues and how they are combined to form the final depth sensation.

1.2.1 Depth cues

The sources of depth information, i.e., depth cues, can be categorized into four groups (Palmer, 1999): pictorial information (e.g., Occlusion, relative size, relative density, height in the visual field), dynamic information (motion parallax and motion perspective), ocular information (convergence and accommodation) and stereoscopic information (binocular disparity).

Pictorial information

Pictorial information can be extracted directly from static and monocular 2D pictures. It also explains why in the case of closing one eye, we can still perceive and judge depth in the real world and why we can perceive good depth even when viewing 2D images.

1) **Occlusion:** occurs when one object hides, or partially hides, another from view. The occluded object is further away than the occluding object.
2) **Relative size:** is the measure of the projected retinal size of objects or textures that are physically similar in size but at different distances. The further away a similar object is located, the smaller the size of the retinal image it produces.
3) **Relative density:** concerns the projected retinal density of a cluster of objects or texture, whose placement is stochastically regular, as they recede into the distance.

4) **Height in the visual field:** are the projected relations of the base of objects in a three-dimensional environment to the viewer, moving from the bottom of the visual field to the top. It yields not only the good ordinal information about distance from the point of observation, but also the potential of absolute distance. The object further away is generally higher in the visual field.

5) **Aerial perspective:** is determined by the relative amount of moisture, pollutants, or both in the atmosphere through which one looks at a scene (E.Cutting and M.Vishton, 1995), When air contains a high degree of either, objects in the distance become bluer, decreased in contrast, or both with respect to objects in the foreground.

Besides the above five monocular depth cues, the way that light reflects from objects provides cues to their depth relationships. Shadows are particularly important in this respect. Thus, **light and shade** (Holliman, 2004a, Seuntiëns, 2006) can be also used as pictorial information. Moreover, **linear perspective** (Holliman, 2004a, Balter et al., 2008) refers to the fact that parallel lines, such as railroad tracks, appear to converge with distance. The more such lines converge, the further away they are.

Most of these monocular depth cues are illustrated in Figure 1-1.

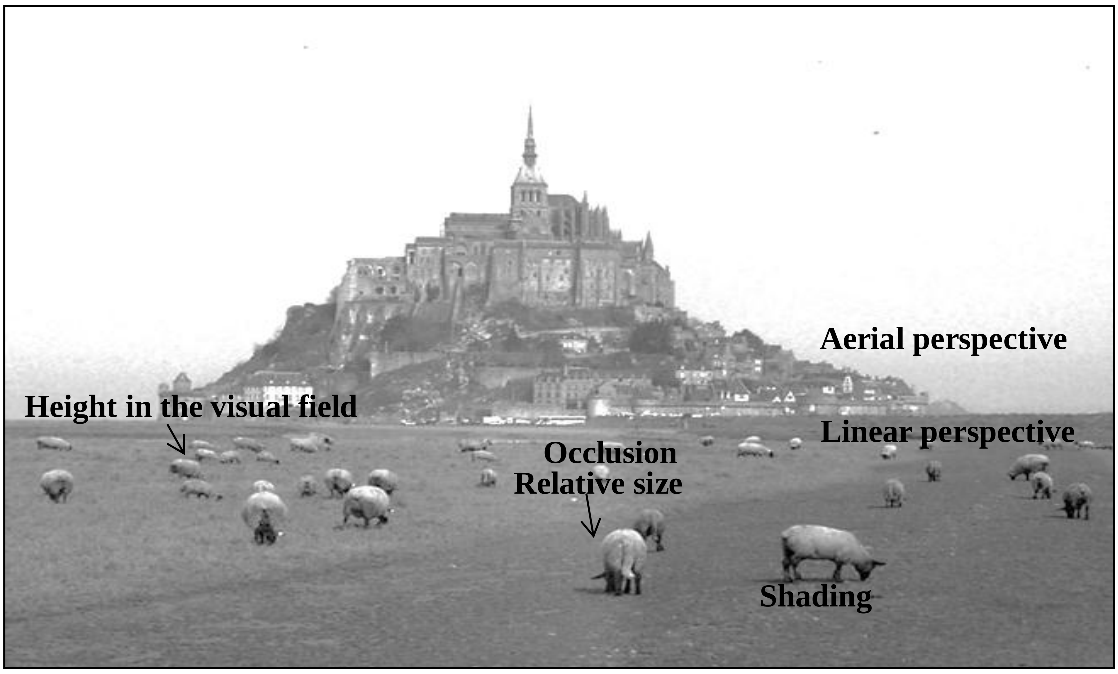


Figure 1-1 : Picture illustrating monocular depth cues in a 2D image (Photographer Jakob Voss)

Dynamic information

Dynamic information occurs when retinal images changes over times because of image motion or head movement.

6) **Motion parallax and motion perspective:** is the relative movement of the projections of several stationary objects caused by observer movement. The motion of a whole field of such objects is called motion perspective. Objects that

are closer will move faster in terms of angular speed than objects that are further away. Figure 1-2 illustrates the motion perspective.

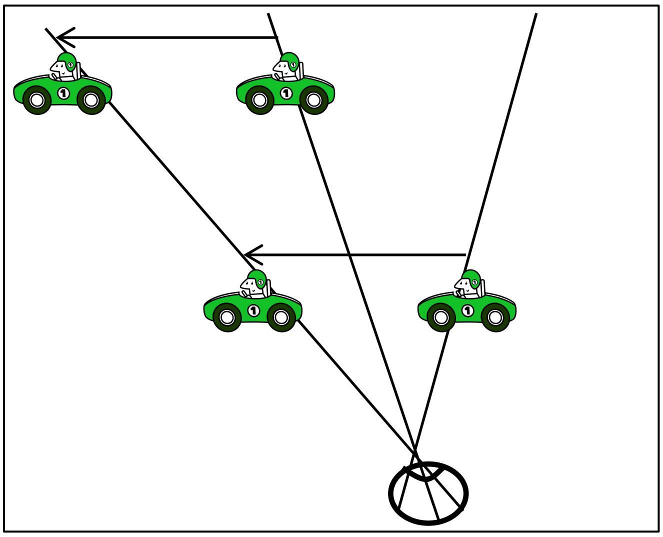


Figure 1-2 : Illustration of motion perspective. A close object that moves the same physical distance as a faraway object will have a larger angular speed, which is a cue of object distance.

Ocular information

The ocular information occurs when the left and right eye balls have relative movement or the lens of the eye change. It consists of two depth cues:

7) **Convergence:** is related to the fixation of the eye and it can be measured by the angle between the optical axes of the two eyes. Fixating on a closer object requires more convergence more than fixating on a distant object. Thus, the convergence level contains the information of the distance between the objects.

8) **Accommodation:** is the change in the shape of the lens of the eye, allowing it to focus on objects near or far while still keeping the retinal image sharp. The muscles of the lens are relaxed when focusing on the objects far away and contracted when focusing on the objects nearby.

The mechanisms of vergence and accommodation system are very complex. The primary stimuli for vergence and accommodation are retinal disparity (Stark et al., 1980) and retinal blur respectively (Phillips and Stark, 1977). However, they are both elicited in response to proximal cues (Hokoda and Ciuffreda, 1983), changes in tonic innervations (Owens and Leibowitz, 1983). Furthermore, vergence and accommodation normally interact and couple with each other (Suryakumar, 2005), i.e., when our eyes fixate on the object of interest, the focus also adapts to guarantee that the perceived image is sharp.

Stereoscopic information

As shown in Figure 1-3, humans have a total field of view (FOV) between 160 to 208 degrees, averaging around 140 degrees for each eye. There is a binocular field of 120 to 180 degrees (Yeh and Silverstein, 1990, Nagata, 1996). In natural vision, when we are looking at a real scene, our two eyes converge on and accommodate the object of interest. Because of the interpupillary distance (IPD), the scene projection on retinal receptors is slightly different for each eye. The human visual system uses these small differences, the binocular disparity, to gain a more accurate judgment of depth and shape (Wheatstone, 1850).

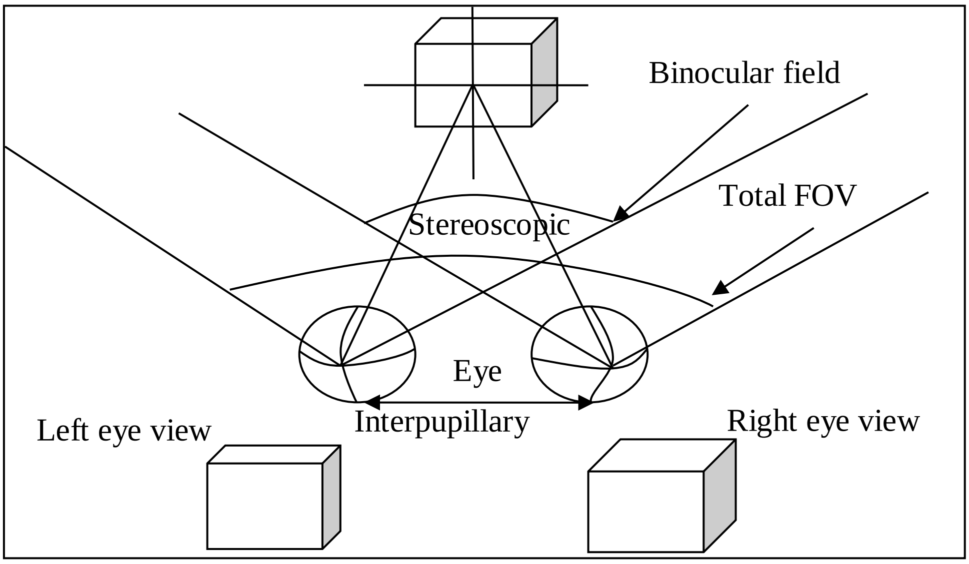


Figure 1-3 : Stereoscopic vision. Retinal images are obtained by geometric projections of the real world. Because of the ocular distance between the eyes, retinal images are slightly different. The visual system can exploit these differences to generate an advanced perception of depth.

9) **Binocular disparity:** due to the fact that human eyes are separated by an interpupillary distance (IPD) of 63mm on average (Dodgson, 2004), each eye receives a slightly different perspective of the same scene as shown in Figure 1-3. The difference in relative position of the projections of the same object on the retinas of the two eyes is called binocular disparity or retinal disparity. The brain can process this disparity information to perceive the relative (perceived distance between objects) and absolute depths (perceived distance from observer to objects). The ability of the brain to process the binocular disparity information is referred to as stereopsis. A more thorough discussion will be given in the next section.

1.2.2 Depth cues and S-3DTV: focus on binocular disparity

Compared with 2DTV, the most important depth cue added by S-3DTV is the binocular disparity. This section focuses on the discussion of how our visual system processes the binocular disparity to generate the 3D sensation. First, we introduce several basic concepts and definitions of functions of the human visual system related to the process of binocular disparity:

- ***Stereopsis:*** is the ability of the brain to process the binocular disparity information in order to generate an enhance depth perception.
- ***Horopter:*** is used to name the geometric arc passing through the fixation point that connects all points in space stimulating corresponding retinal cells (also referred to as cells with zero disparity).

- ***Panum area:*** is defined as the area in space surrounding the horopter where sensory fusion takes place and a single binocular vision is still maintained.

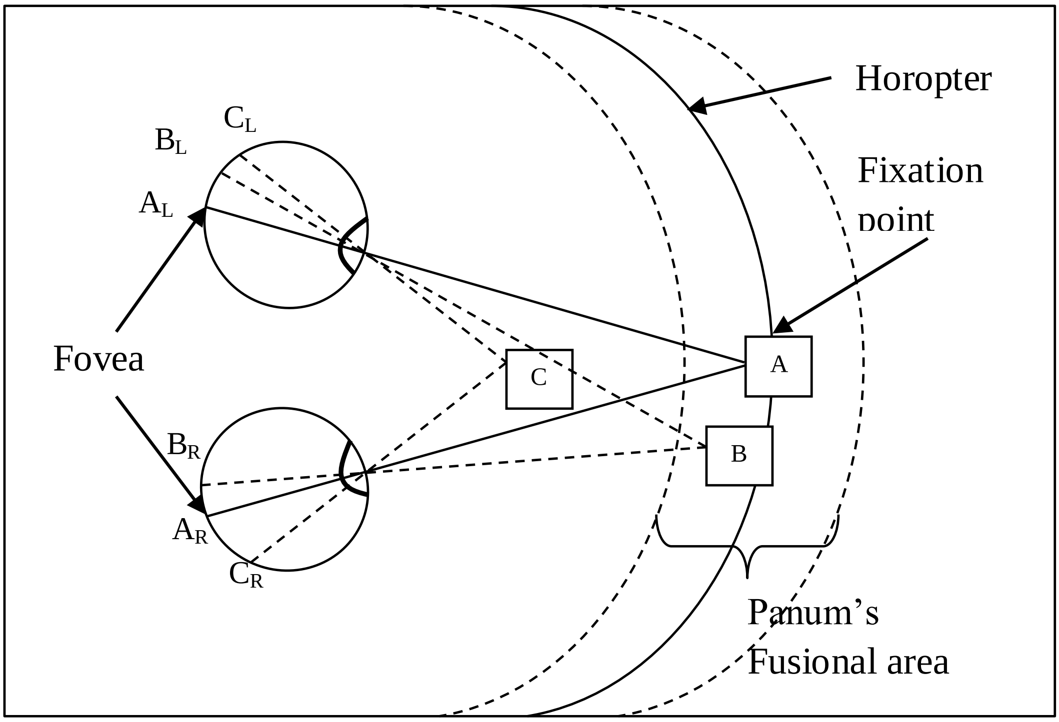


Figure 1-4 : Horopter and Panum's fusional area. The fixation point is on the object A. The fovea is the central part of the retina which is responsible for sharp and color vision. A_L and A_R represent the object A's retinal image in left and right eye, respectively. They are located in the fovea. The horopter often modeled as a circle centered close to the observer's eye and passing through fixation point. Objects on the horopter stimulate corresponding retinal points. The horopter is also referred to as zero retinal disparity region. The binocular disparity of object B can be represented by B_LA_L-B_RA_R. Since it locates in the Panum's Fusional area, the brain can still fuse the image of object B from both eyes. The disparity of object B represents its position in depth to the fixated object A. Object C located outside the Panum area evokes a large crossed disparity. The brain cannot fuse it and thus it will be perceived as diplopia (commonly known as double vision). (Figure adapted from Fig.1 (Patterson, 2007))

When our eyes fixate and focus on an object, it stimulates corresponding retinal areas in both eyes. Thus zero disparity is perceived for the fixated object, as for object A in Figure 1-4. Objects located in front of the horopter generate crossed disparity while object located behind the horopter generated uncrossed disparity. Inside the Panum's fusional area, the retinal images of the object can be fused and a single binocular vision can be maintained. Outside the Panum area, it results in diplopia, as for object C in Figure 1-4. The size of the Panum area or the limit of fusional disparity depends on various factors including eye movements (Yeh and Silverstein, 1990), stimulus properties (Patterson, 2007), temporal modulation of the retinal disparity information (Schor and Tyler, 1981), exposure duration, amount of luminance (Schor and Wood, 1983), and individual difference (Richards, 1970).

Stereoblindness is the inability of perceiving stereoscopic depth using stereopsis. Richards (Richards, 1970) performed a survey among 150 participants and found that 4 % of the participants were unable to use the cue offered by disparity, and another 10 % had great difficulty and incorrectly reported the depth relative to the background. Furthermore, stereoacuity is used to define the ability of human stereopsis to distinguish the minimum disparity. People can detect as low as 2 seconds of arc close to the horopter (IJsselsteijn et al., 2002). Further away from the horopter, stereoacuity reduces. It may vary depending on different condition of the perceived object and environment, e.g., spatial frequency. In, (Schor and Wood, 1983), the authors reported that the stereoacuity threshold depends on the spatial frequency. In their experiments, the stereoacuity threshold increases from 20 second of arc to 5 minute of arc corresponding to the reduction of the spatial frequency from 2-20 cycles per degree to 0.1 cycles per degree. Moreover, stereoacuity also depends on the individual differences. In (Coutant and Westheimer, 1993), B. E. Coutant et al. showed that 97.3 % of people are able to distinguish depth at horizontal disparities of 2.3 minutes of arc or smaller, and at least 80% could distinguish depth at horizontal difference of 30 seconds of arc.

The brain uses binocular disparity to extract depth information from the two-dimensional retinal images in stereopsis, achieving better depth discrimination than only using single image from one eye. Besides better discrimination in depth, the visual acuity can be enhanced by stereopsis when disparities fall within certain boundaries. This is called binocular summation when visual acuity is performed better with two eyes than with one eye (Banton and Levi, 1991). When the differences between the stimuli presented to the two eyes are too large, other precepts become dominant such as binocular mixture, binocular rivalry and suppression, and binocular luster. In case of binocular mixture, the final perceived image is a spatial mix of the left and right images. In case of binocular rivalry, the visibility of the images in both eyes fluctuates: when one eye becomes visible, the view of the other eye is rendered invisible and suppressed. Binocular luster occurs when the luminance or color of uniform areas is different in the two eyes, images will be stable and fused, but shimmer and luster may happen resulting in the failure of depth localization.

1.2.3 Depth cues sensitivity

As presented in previous sections, depth perception arises from a variety of depth cues, e.g., occlusion, binocular disparity, motion perspective, height in the visual field. However, their sensitivity varies depending on the viewing distance. E. Cutting and M. Vishton in (E.Cutting and M.Vishton, 1995) provided thorough results and discussions of the relative information potency of depth cues at various distances (Personal space, Action space and Vista space) as shown in Figure 1-5. The vertical axis, i.e., depth contrast, is defined using distances of two objects, $\boldsymbol{D_1}$ and $\boldsymbol{D_2}$: the ratio of the just-determinable difference in distance between them over their mean distance, $\boldsymbol{2(D_1 - D_2)/(D_1 + D_2)}$. The horizontal axis, i.e., the depth distance, is defined as their mean distance from the observer, $\boldsymbol{(D_1 + D_2)/2}$. A smaller depth contrast value means higher depth sensitivity. The ranking of depth cues by the areas under their curves (see Figure 1-5) within three kinds of spaces are presented in Table 1-1.

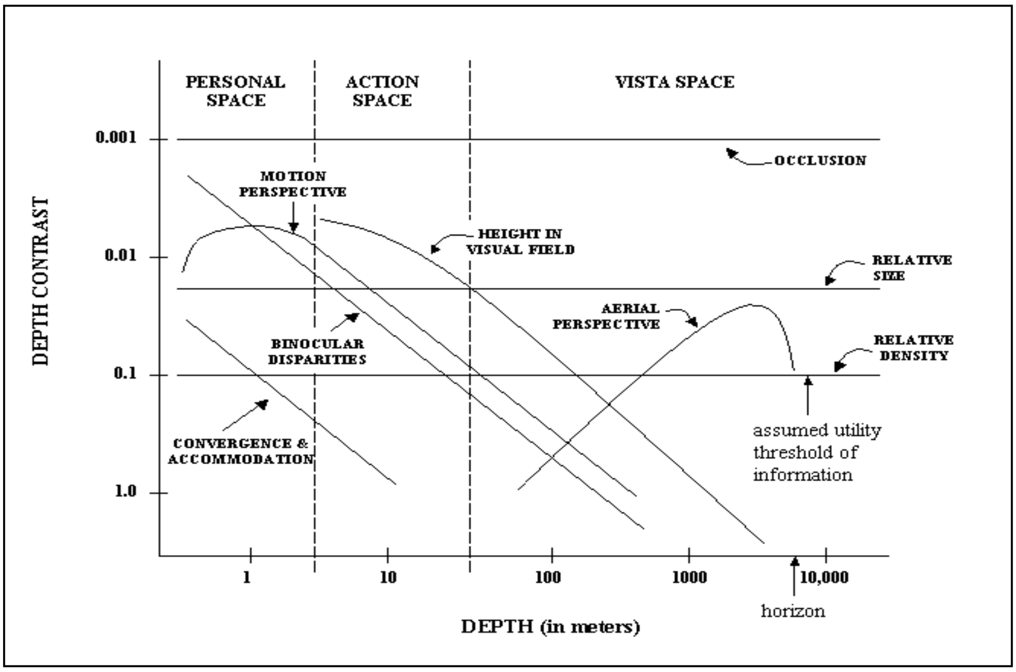


Figure 1-5 : Depth contrast (sensitivity) as a function of the log distance from the observer, from 0.5 to 5000 meters, for nine different sources of information about the layout (E.Cutting and M.Vishton, 1995).

Table 1-1 : Ranking of information sources by the areas under their curves in Figure 1-5 within three kinds of space (E.Cutting and M.Vishton, 1995)

Source of information	Personal space	Action space	Vista space
1. Occlusion	1	1	1
2. Relative size	4	3.5*	2
3. Relative density	7	6	4.5*
4. Height in the visual field		2	3
5. Aerial perspective	8	7	4.5*
6. Motion perspective	3	3.5*	6
7. Convergence	5.5*	8.5*	8.5*
8. Accommodation	5.5*	8.5*	8.5*
9. Binocular disparity	2	5	7

*Float number indicates there are at least two depth cue rated as the same rank.

Starting with pictorial information, occlusion is ranked as the most sensitive depth cue in all three kinds of spaces. For relative size and relative density, their sensitivities are constant with increasing distance. For height in visual field, it is only valid in long distance (Action space and Vista space) and its sensitivity reduces with increasing distance. For relative size and relative density, their sensitivities are constant in all distances. For aerial perspective, it is only valid in vista space. Its sensitivity tends to first increase until a peak and then to reduce when the viewing distance increases. A common trend for pictorial information (expect occlusion) is that their rankings tend to rise with increasing viewing distance.

Concerning dynamic information, motion perspective is ranked the third sensitive depth cues in near distance (Personal space). However, its sensitivity reduces with increasing distance.

Considering ocular information, in long distance (Action space and Vista space), convergence and accommodation are ranked as the least sensitive depth cue. However, in near distance (Personal space), they are still high sensitive depth cues at close distance for specifying the absolute distance of objects (Goodwin, 1995), i.e., the perceived distance from the observer to objects.

For stereoscopic information, the binocular disparity occurs when the object is located in the binocular field as shown in Figure 1-3. It is ranked as second sensitive cue in nine different depth cues in the Personal space in Table 1-1. It indicates that it is a very important and efficient depth cue in close distance. This is also the most important interest and principle behind S-3DTV. Since the viewing distance of S-3DTV is normally located within the personal space (less than 3 meter), adding binocular depth disparity information will provide an important depth cue to the human visual system, thus enhancing the depth perception.

1.3 From binocular vision to stereoscopic imaging system

Inspired by the human binocular vision: "… the mind perceives an object of three dimensions by means of the two dissimilar pictures projected by it on the two retina …", Sir Wheatstone (Wheatstone, 1838) showed that binocular disparity was an effective depth cue by creating the illusion of depth from flat pictures that differed only in horizontal disparity. Thus, he invented the "Stereoscope" in 1838 as shown in Figure I- 1. The basic idea of this device was to separate the left and right viewing channels by additional instruments, e.g., mirrors, and to present different images individually to the left and right eyes. This is also the basic principle and ancestor of modern stereoscopic device.

The original images for the "Stereoscope" were drawings because photography was not yet available. Modern stereoscopic images are captured by two cameras with a horizontal offset. Thus, the simplest modern stereoscopic imaging system for image acquisition and visualization consists of:

- For image acquisition, two video cameras are used to replace the left and right eyes. Binocular disparity information is represented by the slightly horizontal difference between the left and right images, i.e., image disparity.
- For image visualization, the recorded images from the left and right camera are delivered to the left and right eye respectively. Generally, the separation between the left and right image can be carried out using dedicated stereoscopic display technique, for example anaglyph, polarization, active shutter, parallax barrier or

lenticular sheet (Further discussion in Section 1.5.4). In this part, binocular disparity information is visualized on the screen as normally a representation of image disparity, i.e., screen disparity.

Figure 1-6 depicts the principle of such simplest stereoscopic imaging system. Compared with a 2D image system, a stereoscopic imaging system is able to create an illusion of depth sensation by adding binocular disparity information. However, it is important to note that accommodation and convergence information is not reconstructed as the visualization system shows the image information on a planar screen. Moreover, the binocular disparity information may not be identical to viewing the scene directly since it depends on image acquisition parameters and image visualization parameters (detailed in Chapter 3).

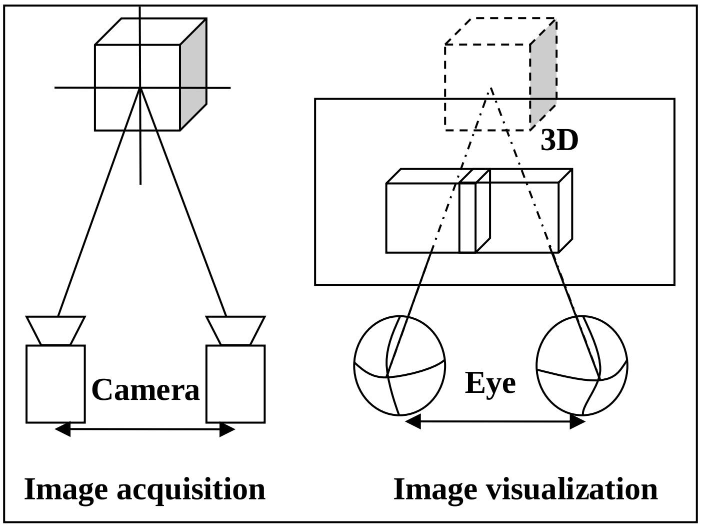


Figure 1-6 : The principle of a simplest stereoscopic imaging system.

On one hand, studies such as (Lambooij et al., 2011, IJsselsteijn et al., 2000) confirmed that the subjective feeling of immersion, depth, naturalness and visual experience are significantly enhanced with stereoscopic images in comparison with 2D ones.

On the other hand, studies such as (Kooi and Toet, 2004, Woods et al., 1993, Yano et al., 2004, Yano et al., 2002, Lambooij et al., 2009a) revealed that problems such as image asymmetry, stereoscopic distortion, decoupling of accommodation and convergence may occur. These negative effects can induce visual discomfort and visual fatigue as detailed in the next section.

1.4 The impact of S-3DTV on visual discomfort and visual fatigue

As presented in the previous section, viewing stereoscopic images is just an illusion of enhanced depth. Compared with viewing 2D images, visual discomfort and visual fatigue are more frequently reported when viewing stereoscopic images. The discrepancies between viewing stereoscopic image and viewing the real scene are recognized as the potential sources of visual discomfort and visual fatigue. In this section, we aim to investigate the potential effect of these discrepancies on visual discomfort and visual fatigue. It will also lead to a general proposal to optimize the stereoscopic images to avoid visual discomfort and visual fatigue.

1.4.1 Definition

In the literature, visual discomfort is used interchangeably with visual fatigue. In this thesis, we make a distinction between them by defining that:

- **Visual discomfort:** it is defined as the observer's particular complaints caused by unnatural visual stimuli. As such, it is a somewhat ambiguous concept, with numerous and widespread causes, symptoms, and associated indicators. It is assumed to be more related to short term effects which can only be explained and measured subjectively.
- **Visual fatigue:** as defined in (Lambooij et al., 2009b), it is a decrease in performance of the visual system. It is assumed to be a subjectively and objectively measurable criterion that is of particular value of ascertaining long-term adaptive processes of the visual system.

Visual discomfort and visual fatigue are not isolated from each other. Perceived visual discomfort measured subjectively is expected to provide indication of the measurable visual fatigue. However, short term visual discomfort may only reflect an adaptation of the visual system and does not cause necessarily visual fatigue. Moreover, the decrease in performance of the visual system is also not always related to visual fatigue. It is essential to distinguish clinically significant visual fatigue from unproblematic, functional adaptation of the visual system.

Concerning the measurements of visual discomfort: Wöpking in (Wöpking, 1992) used subjective assessment of annoyance to measure the visual discomfort on stereoscopic images. Yano et al. in (Yano et al., 2002) used single stimulus continuous quality evaluation (SSCQE) to detect visual discomfort when viewing stereoscopic videos. Jing et al. (Jing Li, 2011) used the pair comparison method in subjective experiment to compare the comfort level between stereoscopic images.

Considering the measurements of visual fatigue: in (Li et al., 2008), electroencephalography (EEG) signal was analyzed to indicate visual fatigue; Yano et al. in (Yano et al., 2002, Yano et al., 2004) suggested that the change of discrepancy of accommodation and vergence may indicate visual fatigue; Emoto et al. in (Emoto et al., 2004) proposed that the change of fusional amplitude and accommodation response is a valid indicator for visual fatigue.

1.4.2 Influencing factors

The main factors, which may cause visual discomfort and visual fatigue when watching S-3DTV, are summarized as follows:

Excessive Screen Disparity

As presented in the last section, screen disparity is a representation of image disparity captured by the left and right camera, leading to the binocular disparity information in the final visualization. Lambooij et al. in (Lambooij et al., 2009b) made a distinction between absolute and relative screen disparity. The absolute screen disparity refers to a disparity offset of the whole retinal image of one eye relative to the other. It may be large and may be overcome by appropriate vergence movement. Yeh and Siliverstein in (Yeh and Silverstein, 1990) demonstrated that with longer stimulus durations and vergence eye movements, fusional retinal disparity can be increased to 4.93 degree for crossed disparity and 1.57 degree for uncrossed disparity. The relative screen disparity refers to the disparity between objects within the retinal images. The single and clear vision can only be perceived as long as the relative disparity remains within the fusion range. Yeh and Silverstein in (Yeh and Silverstein, 1990) also demonstrated that without vergence movement, the fusional disparities are about 27 min of arc for crossed disparity and 24 min of arc for uncrossed disparity.

When relative screen disparities are out of the range of the fusional area, human eyes cannot successfully fuse them and thus diplopia is experienced. Even if the screen disparities are constrained to the fusional range, other problems may be induced by large disparities such as conflicts between accommodation and vergence (Inoue and Ohzu, 1997, Ukai et al., 2009), and conflicts between screen image and reality (Drascic and Milgram, 1996). These effects may potentially cause visual discomfort.

Mismatch of the Accommodation and Vergence

As shown in Figure 1-7, when viewing stereoscopic images, the accommodation plane is assumed to be maintained on the screen plane in order to perceive the images sharply. The converged plane may move out of the screen plane depending on the disparity of the object on the screen. This decoupling of convergence and accommodation is a potential source of visual discomfort and may result in visual fatigue (Yano et al., 2002, Lambooij et al., 2007, Lambooij et al., 2009a, Yano et al., 2004, Emoto et al., 2004).

Many studies reported significant change of accommodation (e.g., the amplitude of accommodation) or vergence function as well as the relationship between accommodation and vergence function (e.g., AC/C ratio, i.e., the change in vergence due to accommodation per change in accommodation in the absence of retinal disparity, or CA/C ratio, i.e., the change in accommodation due to vergence per changes in vergence in the absence of blur) after viewing stereoscopic images. These change have been used as an objective indictor for visual fatigue (Yano et al., 2002, Yano et al., 2004, Emoto et al., 2004, Ukai and Howarth, 2008, Lambooij et al., 2009a). For example, Hiruma et al. (Hiruma et al., 1996) reported an increase in the raise time of accommodation response as well as accommodation error in the stereoscopic performance test. Inoue and Ohzu in (Inoue and Ohzu, 1997) found out that accommodation responses to stereoscopic images differing from viewing a real scene. They concluded that viewing stereoscopic image confuses normal visual functions. Yano in (Yano et al., 2002, Yano et al., 2004) measured the change of accommodation response before and after viewing images in both 2D and 3D conditions. They reported that for some viewers (two in five viewers), the change of accommodation amplitude in the 3D condition is significantly higher than in the 2D condition. Ukai et al. (Ukai and Howarth, 2008, Ukai et al., 2009) reported that viewing stereoscopic image can initiate the changes in the interaction between vergence and accommodation, i.e., altering the AC/A and CA/A ratios.

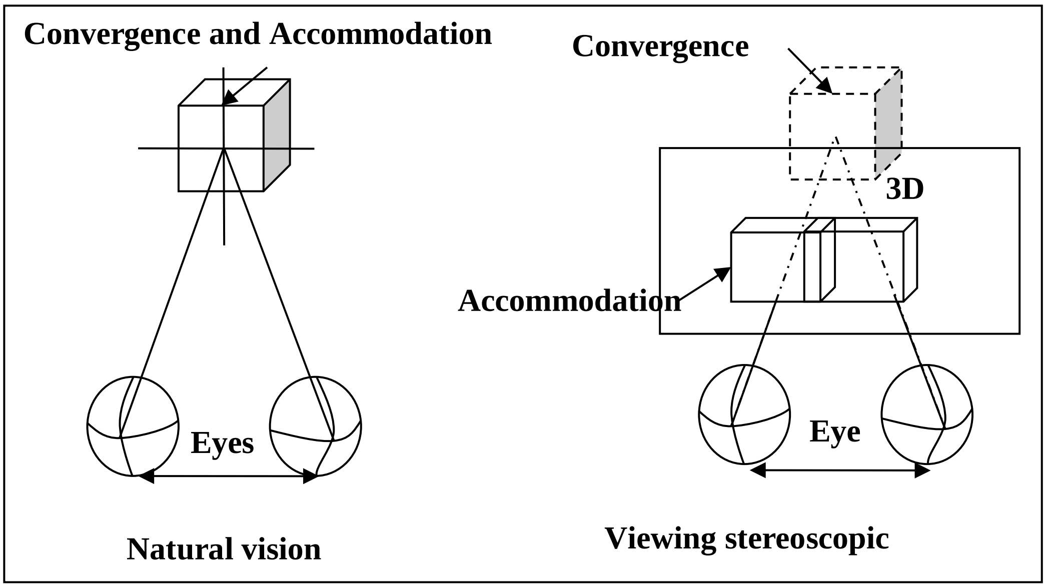


Figure 1-7 : Convergence and accommodation in natural vision and viewing stereoscopic images.

Lambooij et al. in (Lambooij et al., 2009b) clarified that if screen disparity is increased, firstly, vergence movement relocate the retinal disparity within Panum's fusional area and fusional limit is increased. As a consequence, accommodation shifts away from the display under the influence of vergence-driven accommodation.

Depth of focus (DOF) is a lens optics concept that measures the tolerance of placement of the image plane. In the human visual system, the range of DOF is usually used to describe the limits of the accommodative output under natural viewing conditions which concurs with the range of fusion. If this shift of accommodation still remains within DOF, the object of interest is still sharp with regard to focus. However, in case of continuously increasing the screen disparity, the shift of accommodation will be out of the DOF. If defocusing of object occurs, negative accommodation directs accommodation and vergence. Thus, conflict happens. The accommodation-vergence system can cope with certain degree of such conflict, but may operate under stress and thus visual discomfort may increase. If such conflicts keep increasing, three types of erroneous perception can occur: blurred image by loss of accommodation, diplopia by loss of the convergence, or both.

The comfortable viewing zone describes a range in depth where objects can be reconstructed on a planar screen without inducing visual discomfort. The image located inside the comfortable viewing zone remains sharp and can be fused without decoupling of accommodation and convergence.

Many studies have proposed thresholds for disparity to guarantee comfortable viewing. These thresholds are indicators of the comfortable viewing zone. The traditional rule-of-thumb threshold for disparity is a maximum of 70 minutes of arc. This value was computed from the human eye's aperture and depth of focus. Wöpking in (Wöpking, 1992) confirmed this value by subjective experiment on stimuli in a wide range of disparity (0-140 minute of arc) and depth of focus condition. Kooi et al. (Kooi and Toet, 2004) conducted a subjective test for visual comfort. They proposed that the maximum horizontal disparity should be maintain between 2 and 3 PD (1 prismatic diopter = 0.57°). In (Lambooij et al., 2007, Lambooij et al., 2009b), Lambooij et al. defined the accommodation threshold as the limit of depth of focus (0.3 diopters, diopter is a reciprocal value of distance) and vergence system threshold as "the zone of clear, single binocular vision" (1° of disparity). They reported that these two thresholds resemble each other. Consequently, 1 ° for disparity was proposed as a general threshold. In ITU-R BT.1438 Recommendation (ITU, 2000) for subjective quality assessment of stereoscopic television pictures, a threshold of ±0.3 diopters was recommended as the desirable range of depth. A more conservative threshold of ±0.2 diopters was proposed in (Yano et al., 2002, Yano et al., 2004).

The above limit for the comfortable viewing zone does not take into consideration of the movement of objects in images. Yano et al. in (Yano et al., 2004) reported that visual discomfort can still be induced if the images were moved in depth according to a step pulse function within a the comfortable viewing zone of ±0.2 diopters. Jing Li et al in (Jing Li, 2011) gave another indication that a stimulus which has small relative disparity but fast velocity might have a similar effect on visual discomfort compared to a stimulus which has large relative disparity and slow velocity. The above findings might indicate that the range of the comfortable viewing zone should be reduced in case of faster movement. In this thesis, several studies aim to determine and provide specification for the comfortable viewing zone (see Chapter 5 and Chapter 6)

Stereoscopic geometrical distortion

Stereoscopic geometrical distortion is mainly related to the distortion in depth perception. There are two well-known phenomena related to stereoscopic geometrical distortion: the puppet-theatre effect and the cardboard effect. The puppet-theatre effect makes a three dimensional image (3-D) look unnaturally small compared with the real object; The cardboard effect refers to the phenomenon in which the observers of stereoscopic images get the impression that individual objects in the images are flattened like a cardboard (i.e.,, not 3D), although they appear with correct perspective (Yamanoue et al., 2006). Boev et al. (Boev et al., 2009) explained that the puppet theatre effect is caused by inconsistency between the binocular and perspective depth cues and cardboard effect is mainly due to the limited depth or disparity information or coarse depth quantization.

Woods et al. in (Woods et al., 1993) established a geometry model of stereoscopic camera and display system and demonstrated that depth-plane curvature, depth non-linearity, depth and size magnification can be caused by an inappropriate choice of camera parameters and display system parameters. Yamanoue et al. in (Yamanoue et al., 1998) presented the orthostereoscopic condition for 3D HDTV which can avoid stereoscopic geometrical distortion. Their subjective evaluation confirmed that under this condition, the images look more natural than in other conditions. However, it requires very strict shooting and visualization conditions. The same authors in (Yamanoue, 2006, Yamanoue et al., 2006) focused on the comparison of the parallel and toe-in camera settings. They reported that parallel camera setting (see chapter 1.5) can avoid puppet-theatre effect. Most of the above mentioned studies focused on geometry prediction of the stereoscopic depth distortion and only very limited works investigated its perceptual impact on visual comfort.

View asymmetry

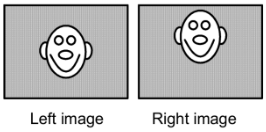
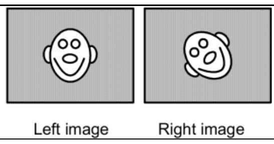
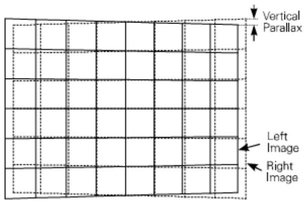
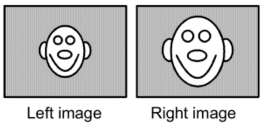
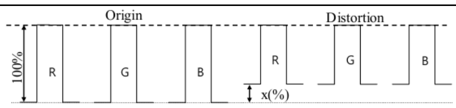
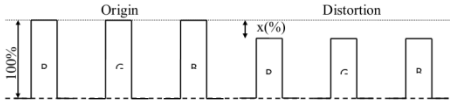
Stereoscopic 3D content contains two views, i.e., the left view and the right view. View adjustments including color, brightness, temporal sampling and geometry calibration are very important. View asymmetries can be induced by many reasons, e.g., imperfect filters and lenses or misalignment of optics. There are various types of view asymmetries:

- Geometrical asymmetries: include vertical shifts, rotation of one view and keystone distortion. They can be induced by geometry errors from image acquisition (e.g., misalignment of the left and right camera in a stereoscopic two camera system) or image visualization (e.g., misalignment of left and right projectors in a two projectors based stereoscopic display).
- Optical asymmetries: are mainly related to the differences of focal length. Blur and magnification in one view will be induced by the differences of focal length.
- Luminance asymmetries: can be induced by the imperfect of filters or desynchronization of white level, black level and color gamut in stereoscopic camera system.
- Color asymmetries: can be induced by imperfect filter in the camera such as semi-transparent mirrors used in order to reduce the stereoscopic base or specific 3D visualization technique such as anaglyph glasses.
- Ghost image or crosstalk: Imperfect separation of the left and right view in stereoscopic image system makes a small proportion of one eye’s image perceptible to the other eye. This phenomenon is known as crosstalk or image ghosting.
- Temporal asymmetries: the desynchronization of 3D capture or visualization system especially active shutter glasses technique can induce temporal asymmetries. This can cause contradiction among psychological depth cues resulting in an increase of visual discomfort.

Kooi et al. in (Kooi and Toet, 2004) conducted a subjective experiment of visual comfort to assess a wide range of view asymmetries. Their results showed that nearly all binocular image asymmetries seriously reduce visual comfort if they are presented in large enough amounts. Balter et al. in (Balter et al., 2008) summarized the visibility thresholds for different types of view asymmetries in the literature. They noted that even in case of degradations remaining under visibility thresholds, cumulative effect of asymmetries can induce an increase of visual fatigue. It is therefore recommended to reduce them to a minimum.

Table 1-2 illustrated the stereoscopic asymmetries and their thresholds (visibility threshold and visual discomfort threshold) proposed in the literature. All the visibility thresholds and visual discomfort thresholds were derived from subjective experiments (Kooi and Toet, 2004, Fournier, 1995a, Seuntiëns et al., 2005, Ikeda and Nakashima, 1980, Ion-Paul and Hanna, 1990). Visibility thresholds are more critical than visual discomfort threshold in Table 1-2.

Table 1-2 : Illustration and threshold (visibility and visual discomfort) of stereoscopic asymmetries, adapted from Table 6-1 page 55 in (Balter et al., 2008)

Asymmetry	Illustration	Threshold
Geometry: Vertical shift of one view		34 minute of arc (1 PD) (C) [a] (Kooi and Toet, 2004)
Geometry: Rotation of one view		0.5 °(V) [a] (Fournier, 1995b)
Geometry: Keystone distortion		3 minutes of arc (V) (Ion-Paul and Hanna, 1990) / 1PD (0.57°) (C) (Kooi and Toet, 2004)
Optical aspect: Focal length difference		1%[b] (V) (Fournier, 1995b) and 2.5%[c] (C) (Kooi and Toet, 2004)
Optical aspect: Definition difference	Difference of image definition between views	30% [b] (V) (Fournier, 1995b)
Luminance: Black level difference		1%(0.1 dB) (V) (Fournier, 1995b)
Luminance: White level difference		15%(1.5 dB) (V) (Fournier, 1995b)
Ghost image or crosstalk	Perception of a proportion of the right image on the left eye and/or the left image on the right eye	From 0.2% to 7% (V) (Fournier, 1995b) / 5% (C) (Kooi and Toet, 2004) / 2% (V) (Seuntiëns et al., 2005)
Color difference	Colorimetric of left and right images are different	From 15 to 100 nm in wavelength limit (V) (Ikeda and Nakashima, 1980)
Temporal: Synchronization difference	Acquisition and/or restitution of left and right images not at the same instant	2 frame (25 frames per second) difference leads to a significant quality drop (Goldmann et al., 2010b)

[a] C for visual discomfort , V for visibility
[b] experiment setup: SD resolution, 4.5 times image height viewing distance
[c] experiment setup: 1024x768 resolution, 170x128 cm^2 screen size, 185cm viewing distance

Stereoanomaly

Stereoanomaly is the failure to see difference in depth when the viewer is presented with stimuli having different magnitudes of stereoscopic disparity (van Ee and Richards, 2002). For example, when watching a stereoscopic image, certain individuals may perceive the crossed disparity (which should be the front depth relative to the horopter) as the back depth or the uncrossed disparity (which should be the back depth) as the front depth. Patterson in (Patterson, 2007) reported stereoanomaly can occur in about 20-30% of people under degraded stimulus and it may induced visual discomfort and visual fatigue. The way to avoid stereoanomaly is to present the image under non-degraded conditions (enough luminance and resolution) or to enhance the disparity information with other depth or distance cues.

Windows violation

Windows violation (Mendiburu, 2009) is a cognition level depth cue conflict. It is a well-known conflict in the 3D film production industry. It occurs when crossed disparity objects perceived in front of the display window are cut off by the screen border. It is physically impossible that a window frame, seemingly appearing behind the object, is able to obscure it. Human brain may be confused and not able to process this conflict so that visual discomfort will be induced.

1.4.3 Discussion

In summary, to avoid visual discomfort and visual fatigue, stereoscopic images should fulfill the below requirements:

- To be presented within the comfortable viewing zone in depth to avoid excessive screen disparity and decoupling of accommodation and convergence
- To adapt the camera parameters considering the visualization environment to avoid stereoscopic distortion
- To avoid the image asymmetries or at least to guarantee that the image asymmetries level are lower than the perceptual thresholds
- To present the image in non-degraded conditions to avoid stereoanomaly
- To design the scene by considering the final depth rendering in order to avoid windows violation

1.5 QoE issues in modern S-3DTV broadcast chain

In the last section, we presented and discussed the impact of S-3DTV on QoE (focusing on visual discomfort and visual fatigue) based on a simplified stereoscopic imaging system. In the television broadcasting domain, a stereoscopic imaging system is represented by a S-3DTV broadcast chain. The S-3DTV broadcast chain is an end to end solution, including production and 3D representation format, coding, transmission, visualization terminal and finally the end-user’s perception as presented in Figure 1-8. In different parts of the S-3DTV broadcast chain, various techniques are available. In this section, we aim to explore the QoE issues of different techniques in the S-3DTV broadcast chain.

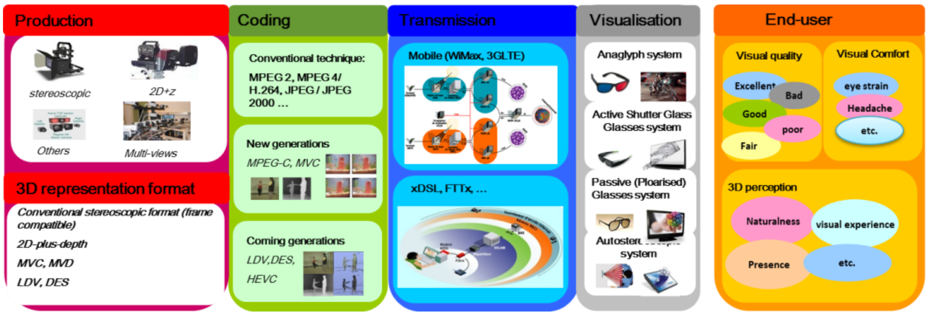


Figure 1-8 : 3DTV broadcasting chain

The advantages and drawbacks of different technologies in different steps of the broadcast chain will be reviewed in order to understand their potential impacts on the final QoE as well as to ease the selection of an optimized stereoscopic imaging system.

1.5.1 Content production

Compared with 2D image production, a stereoscopic 3D image production requires additional information, i.e., the binocular depth information. There are different 3D production systems to capture and generate 3D stereoscopic images, e.g., monoscopic systems with 2D to 3D conversion (automatic or semi-automatic conversion), monoscopic systems with additional depth sensor, the traditional stereoscopic two-camera systems (mirror systems, Side-by-Side rig systems and etc.), multi-view systems (more than two cameras, with or without additional depth sensor) and synthetic content production.

Monoscopic systems with 2D to 3D conversion

These systems are identical to traditional monoscopic systems. They do not require any additional equipment to capture the depth information. However, specific algorithms are required to extract the depth information from the 2D images. Various algorithms were proposed to extract the depth maps from monocular depth cues, such as defocus (Ziou and Deschenes, 2001), linear perspective (Battiatoa et al., 2004), atmosphere perspective (Cozman and Krotkov, 1997), shading (Ruo et al., 1999), relative size (Loh and Hartley, 2005), height in the visual field (Jung et al., 2009) and occlusion (Redert, 2005). Wei et al. in (Wei, 2005) investigated the existing 2D to 3D conversion algorithms developed in the past 30 years. They concluded that a single solution to convert the entire class of 2D images to 3D images does not exist. The authors stated that no one depth cue is superb or indispensable for depth perception so that it is necessary to combine the suitable depth cues in order to achieve a robust all-round conversion algorithm. Machine learning algorithms such as (Saxena et al., 2005) were proposed as a new and promising research direction for 2D and 3D conversion.

Even though the accuracy of depth maps can be guaranteed by advanced algorithms using different monocular depth cues, how to reconstruct the stereoscopic views (i.e., the left and right views) especially the occlusion layer is still a challenge.

In movie industry, semi-automatic 2D to 3D conversion algorithms (e.g., (Chen et al., 2011)) are still widely used in order to guarantee the conversion quality. However, human intervention increases the production time and expense.

Monoscopic systems with additional depth sensor

These systems usually combine traditional monoscopic camera and an additional depth sensor to capture one 2D image and one depth map image simultaneously. For example, the ZCam system (Iddan and Yahav, 2001) as shown in Figure 1-9 consists of a RGB camera and a Depth camera which both share the same optic channel. A laser ring illuminator is installed around the head of the focus length. The sensor of the depth camera can recorded the reflected laser lights from the scene which contain depth information of the scene.

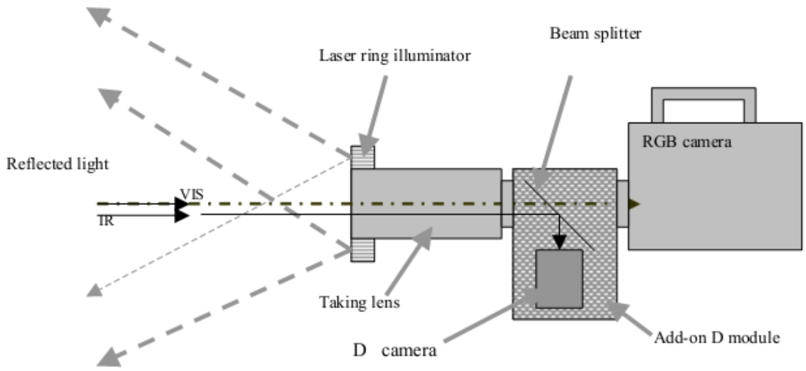


Figure 1-9 : Monoscopic camera + depth sensor, ZCam system (Fig. 7 in (Iddan and Yahav, 2001))

The range of the depth sensor depends on the strength of the laser ring illuminator. Normally, it is within 10 meters. The advantage of this kind of systems is that the depth maps is associated to 2D images and normally it is a monochrome 8 bits image which can be compressed and stored like a conventional 2D image (Fehn, 2001, Fehn, 2003). The drawback for these systems is that: 1) due to the limit (range and luminance) of the depth sensor, it may not be appropriate for outdoor content shooting; 2) the quality of reconstructed stereoscopic views may be another issue since occlusion layers are still not recorded.

Stereoscopic two-camera systems

Stereoscopic two-camera systems use two dedicated 2D cameras (representing the left and right eye as shown in Figure 1-6) positioned at slightly different viewpoints in the same scene to capture the stereoscopic images.

There are two types of camera configurations for stereoscopic two-camera systems: toed-in (converged) and parallel as shown in Figure 1-10.

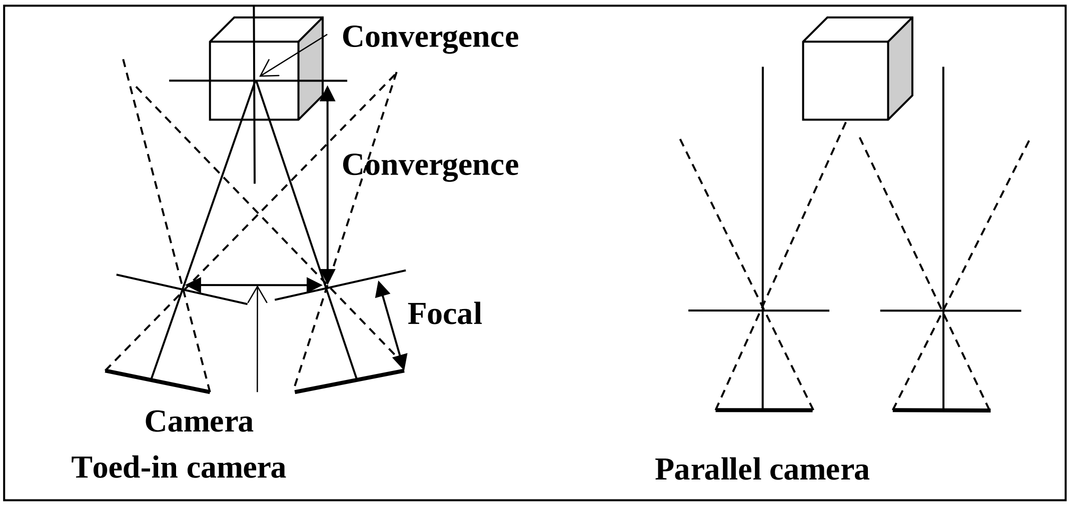


Figure 1-10 : Toed-in camera (left) and parallel camera (right) configurations

In the toed-in camera configuration, the optic axes of the left and right camera are crossed in a convergence point (e.g., object of interest). In the parallel camera configuration, the optic axes of the two cameras are parallel, or in another words, converged on an infinite point. Woods et al. in (Woods et al., 1993) analyzed the geometry of stereoscopic image system and recommended that the parallel camera configuration is used in preference to the toed-in camera configuration since parallel camera configuration can eliminate keystone distortion and depth plane curvature. Yamanoue et al. (Yamanoue, 2006) demonstrated that parallel setting can maintain linearity during the conversion from real space to stereoscopic images, however, the toed-in setting cannot. The same authors in (Yamanoue et al., 2006) conducted a subjective experiment to compare the impact of camera settings on the puppet theatre and cardboard effects. They demonstrated that toed-in camera may produce puppet theatre effect while parallel camera may not, and both camera settings may produce cardboard effect. Furthermore, many studies (Woods et al., 1993, Holliman, 2004b, IJsselsteijn et al., 2000, Goldmann et al., 2010c) also demonstrated that the camera parameters, such as camera baseline, focal length, convergence distance, affect the final depth perception of stereoscopic images. These parameters should be carefully defined in order to avoid stereoscopic distortion.

For modern stereoscopic two-camera systems, some additional device can be added to facilitate the image acquisition. For example, a mirror rig can be used to reduce the camera baseline to small values in case of large cameras like broadcast-quality ones.

The main advantage of these camera systems is that they capture the stereoscopic images (the left and right view) directly without requiring any reconstruction or conversion for final visualization. The potential QoE issues are: 1) the camera configuration and shooting parameters affect the final depth perception. They should be carefully selected and defined by considering the scene parameters (e.g., depth budget) and the visualization parameters (e.g., viewing distance and screen size); 2) the calibration including positioning, luminance, colors are necessary to avoid image asymmetries (see Table 1-2).

Multiview systems

These systems are camera array composed of more than two traditional monoscopic cameras. The advanced model of these systems can also add multi depth sensors (Jolly et al., 2009). Free viewpoint television (FTV) is a typical application example for such a system. Tanimoto in (Tanimoto, 2006) introduced a real-time FTV system which was composed of 100 cameras as shown in Figure 1-11. Kubota et al. in (Kubota et al., 2007) gave a survey of multiview imaging on 3DTV and stated the “3DTV and FTV are not mutually exclusive. On the contrary, they can be very well combined with single systems as they are both based on a suitable 3-D scene representation”.

Figure 1-11 : A **100-cameras multiview system** (Fig 6. from (Jolly et al., 2009))

In general, multiview systems with a larger number of cameras can provide a more precise 3-D representation, resulting in higher quality views through the rendering and display process, and vice versa. However, there are still numerous challenges: 1) capturing and storing a large number of images in real time require very high performance and large capacity of transmission and storage system; 2) accurate calibration of camera positions, luminance, color and optics are required.

Synthetic content production

Thanks to the rapid development of 3D computer graphic techniques (Watt, 1999, Buss, 2003), producing 3D synthetic content only required virtual cameras in the 3D synthetic scene in order to capture the stereoscopic views or generate the depth maps. The great advantages of 3D synthetic content production are 1) camera position and calibration can be precisely controlled; 2) information of the 3D scene can be easily extracted and manipulated. Thus, the accuracy of depth map is no longer a problem.

Concerning the QoE issues of the aforementioned stereoscopic production systems, several points can be drawn as follows:

- Concerning depth map based systems such as 2D to 3D conversion and monoscopic system with additional depth sensor, the native lack of occlusion layer information and the precision of depth map may affect the final reconstructed stereoscopic images' quality;
- Considering stereoscopic two-camera systems, camera configuration and shooting parameters may affect the final depth perception. Moreover, calibration of cameras is very important to avoid image asymmetries;
- Multiview systems seem to be able to reconstruct the most precise 3D information, however, the calibration between the cameras is even more complex;
- 3D computer graphic production may provide excellent quality of each view. However, it only can be used for synthetic scenes.

1.5.2 3D representation format

There are various 3D representation formats (Macchiarella, 2010, Gautier et al., 2010) available in the literature, such as conventional stereo video format, 2D-plus-depth-format, multi view video format and multiview video plus depth format, layer depth video format and depth-enhanced stereo format. In this section, the merits of each format are discussed along with the drawbacks and limitations.

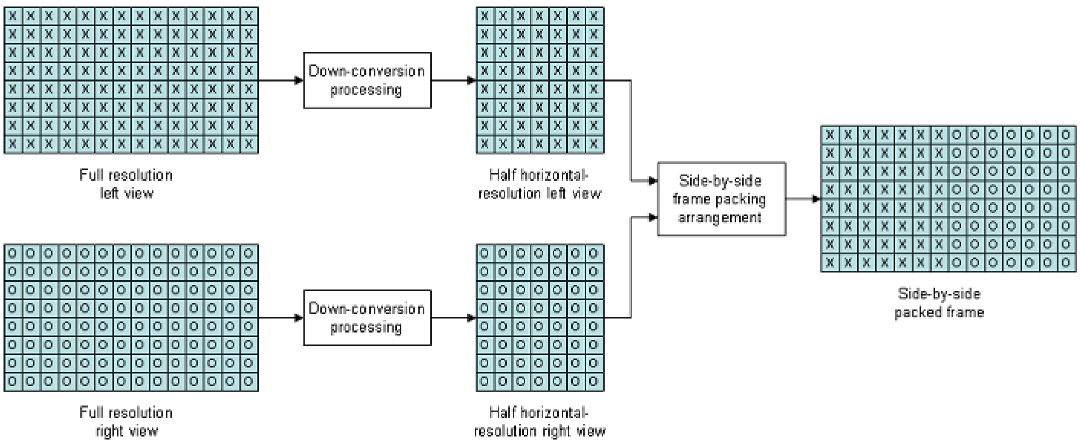
Conventional stereo video format

Conventional stereo video format consists of a pair of sequences, showing the same scene for the left and right view. Normally, double capacities are required to transmit and store such data. In order to be compatible to conventional broadcast chain and network, frame compatible formats such as Top-and-Bottom and Side-by-Side were proposed in HDMI specification(HDMI, 2009) and DVB Document A154(DVB, 2011). As shown in Figure 1-12, horizontal sub-sampling and vertical sub-sampling are implemented to generate Side-by-Side and Top-and-Bottom formats, respectively. The resolution per view are halved, thus, it is compatible for conventional HD video frame.

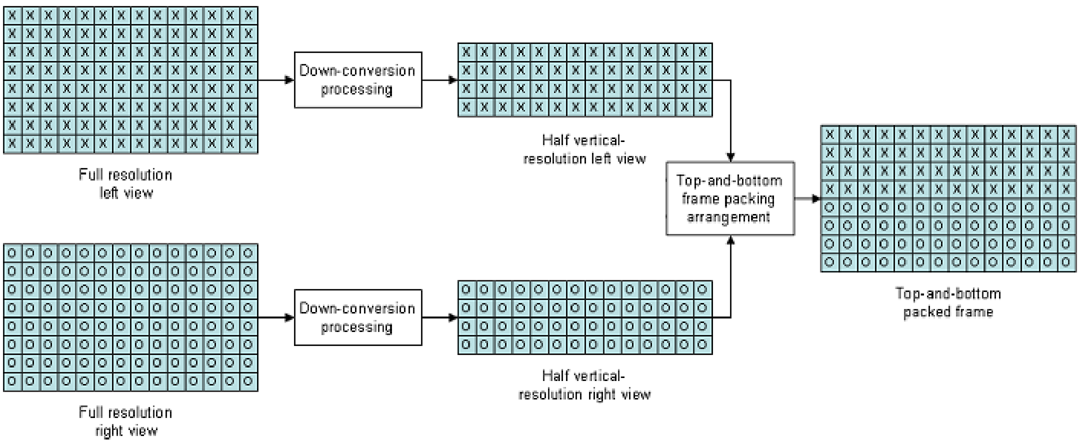
2D-plus-depth format

2D-plus-depth format comprises additional depth information with every 2D image. It is described in a Philips' white paper (Solutions, 2008) and MPEG-C part 3 (ISO, 2007). Instead of transmitting a two view color video as conventional stereo video format, the 2D-plus-depth format consists of a single view image and an associated depth map as shown in Figure 1-13.

The 2D-plus-depth format is not totally equivalent to a pair of stereo images because occlusion information is not contained in the depth map, involving the apparition of holes when novel views are generated. If the navigation is small around the original images, these holes can be filled by padding pixels or with inpainting technique (Jantet et al., 2009). However, if the occlusion layer is too large or too complex, visual artifacts may occur due to the limit of inpainting algorithms.



Side-by-Side



Top-and-Bottom

Figure 1-12 : Side-by-Side and Top-and-Bottom frame compatible formats (adapted from Fig 8 and 10 in (DVB, 2011))

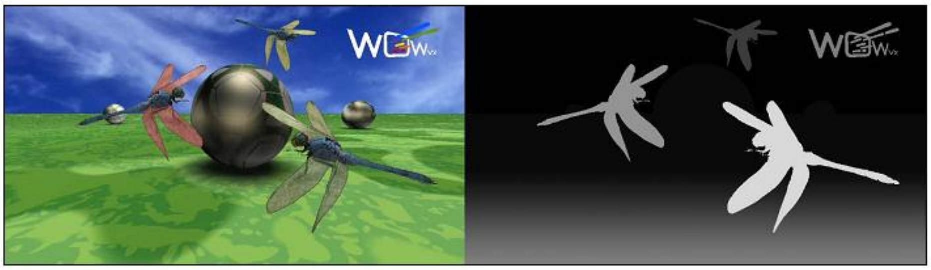


Figure 1-13 : 2D-plus-depth format (Fig. 1 from (Solutions, 2008))

Multi View Video format and Multi View plus Depth format

Multi view video format (MVV) (Flierl and Girod, 2007) comprises a number for views capturing a scene from different viewpoints. Having N views form slightly different viewpoints allows for a 3D impression within a range by presenting two adjacent of the N views as a stereo pair to the user. The drawback of this format is that it requires a huge capacity to store and transmit the data of a number of views. Multi view plus depth format (MVD) is an advanced format, consisting of multi views of 2D-plus-depth information. Since a number of novel views can be reconstructed by 2D-plus-depth information, it requires less capacity than MVV format.

Layered Depth Video format and Depth-Enhanced Stereo format

To cope with the problem of lack of occlusion layer of the 2D-plus-depth format and the problem of huge capacity requirement of MVV and MVD formats, Layered depth imaging (Shade et al., 1998) representation format was proposed in MPEC-C part 3 as an extension of 2D-plus-depth format. It consists of representing color and associate depth in pixels in their consecutive position along some depth layer. This extension is also called the “Declipse format” in Philips’ white paper (Solutions, 2008) and layered depth video (LDV) in 3D4YOU project (Kerbiriou et al., 2010). Figure 1-14 gives an example of the LDV format. The top two images are color images and the bottom images are depth images while the left two images are main layer images (as the 2D-plus-deph format) and occlusion layer images. Compared with 2D-plus-deph format, the additional occlusion layer images can facilitate the generation of new views.

Figure 1-14 : LDV format: **color (top) and depth (bottom) of main layer (left), occlusion layer (right) (Fig. 5-1, page 52 from (Kerbiriou et al., 2010))**

Bruls et al. in (Bruls et al., 2007) proposed LDV-R format, which further adds a second color view (or the right view) to a classical LDV format in order to improve the quality of reconstructed novel views. Smolic et al. in (Smolic et al., 2009) proposed Depth Enhanced Stereo (DES) as shown in Figure 1-15. This format extends the conventional stereo video format with the LDV capabilities. It provides the stereo backward compatibility and also enables depth-based view synthesis with an improved quality compared with single view LDV.

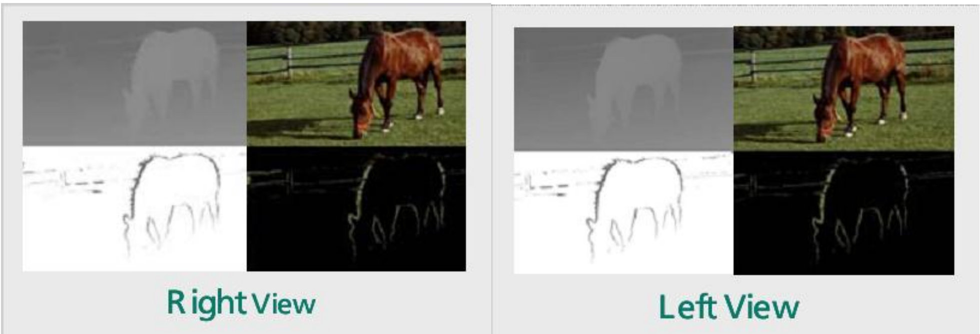


Figure 1-15 : Depth enhanced stereo format (Fig.7 from (Smolic et al., 2009))

The QoE issues of the 3D representation formats related to current challenges are:

- For frame compatible format, the resolution reduction effect may affect the quality and requires further investigation;
- For depth-map based format, even for LDV format, the quality of the reconstructed novel view is still not comparable to native stereo views (Kerbiriou et al., 2010).

1.5.3 Coding and transmission

Today, as suggested in the DVB Bluebook A154 (DVB, 2011), current S-3DTV digital video broadcast tends to re-use conventional codecs (e.g., MPEG-2 or H.264/AVC) and the original HD channel, to compress and transmit stereo video signals in frame compatible formats such as Side-by-Side and Top-and-Bottom. However, the view resolution reduction effect and 3D coding artifacts (IJsselsteijn et al., 2002) in frame-compatible format may have potential negative influence on the image quality. In order to highlight the value of 3D video service, the 3D images should maintain the same texture quality as the 2D HD images. To ensure that the artifacts are the same in both the left and right channels, the only practical way is to use higher compression bitrate than typically used for HD (moote and lennon, 2010).

For conventional stereoscopic format with full definition, if using simulcast coding scheme, double HD bandwidths are required. Although the coding efficiency can be improved by advanced coding method (see Appendix A: 3d video encoding) such as Multi view coding (MVC) (Merkle et al., 2007b, Smolic et al., 2007, Merkle et al., 2007a), it still requires higher capacities than conventional HD channels. In (Yamagishi et al., 2011), the author showed that the quality and depth perception of the full spatial resolution video sequence are higher than those of side-by-side video sequence for uncompressed video sequences. For compressed video sequences, they recommended that for full spatial resolution format encoded using MVC (JMVC version 8.3), 9 Mbps was required to maintain high image quality (above “fair” MOS) and 5 Mbps was required to maintain high depth perception (above “fair” MOS).

2D-plus-depth format is claimed to be able to save the bitrate due to the fact that depth maps are more coding-friendly than color texture images. However, it still might add at least 20% to 30% to the HD bitrates (Fehn, 2003). For MVV format or MVD format with MVC coding scheme, more bandwidths can be required depending on the amount of views. For LDV format, due to immaturity of the specific coding scheme, the studies in (Kerbiriou et al., 2010) showed it even required more bitrates (up to 60%) to maintain the same PSNR as the MVD with 2 views. Mixed resolution coding (Brust et al., 2009) or asymmetric coding (Seuntiëns et al., 2006, Kalva et al., 2007, Saygili et al., 2011) may be an alternative solution to save bitrates especially for bandwidth constrained application (e.g., mobile network). However, the view asymmetries artifacts from mixed resolution coding may have potential impact on visual discomfort and visual fatigue (Kooi and Toet, 2004). Overall, advanced 3D video services require larger bandwidth for signals transmission.

Besides the bandwidth requirement, Bing in (Bing, 2010) emphasized that network transmission protocol, IPTV protocol, Quality of Service (QoS) and Quality of experience (QoE) algorithm or protocol, error concealment methods should also be optimized for 3D video networking.

In summary, the QoE issues in coding and transmission are:

- Considering coding and transmission of frame compatible formats, it may require higher bitrate than conventional 2D-HD channels to ensure the same level of texture quality as 2D HD images.
- Concerning full definition 3D or advanced 3D representation formats such as 2D-plus-depth, LDV, they increase computation complexity for coding and require higher bitrate for transmission.
- Mixed resolution coding may induce view asymmetries with potential impact on visual discomfort and visual fatigue.

1.5.4 Visualization terminal

The advanced real 3D technique such as volumetric and holographic (Alatan et al., 2007) are still far away from maturity for 3DTV application. In this section, we concentrated on the 3D display techniques which are currently available in the market.

In (Holliman, 2004a) and (Dodgson, 2005), the authors gave an introduction of different display technologies. Here, we categorize those technologies by whether the additional glasses are required for separation of views.

The most common visualization systems using glasses are the followings:

- **Anaglyph technique**: anaglyphic stereo images are stereo pairs of images in which each image is shown using a different color. The two images are overlapped and then watched using glasses with corresponding color filter. Common association of colored filters used for TV application is red/cyan and yellow/blue. The advantage of common anaglyph technique is the backward compatibility to existing TV displays. The drawback is the native color asymmetries can cause visual discomfort and even visual fatigue. Advanced anaglyph technique (Froehly et al., 2003) uses specific filter to multiplex the primary colors into different wavelength. Thus, color asymmetries problem can be reduced. However, it is not compatible with existing TV display.

- **Polarized technique**: specific filters such as orthogonal polarizing filters (linear polarizing) or circular polarizing filters are used to separate the left and right views in the visualization system. User has to wear glasses with the same polarized filters to view the left and right image on the left and right eye, respectively. Both linear polarizing and circular polarizing technique are sensitive to the viewing position. If the filters in the screen and the filter in the glasses are not aligned, luminance reduction and crosstalk can be induced. Compared with the linear polarizing technique, circular filter allows viewers to lean their head to the left and right directions.

  The advantage of polarized technique is that it can provide full color image. Considering a two-projector based solution such as in the cinema case, it can also provide two full spatial and temporal resolution images. For 3D television application, there are two types of displays with polarized solution: line interleaved displays and column interleaved displays. For line interleaved displays, odd and even lines with different polarized filter, represent the left and right views (or right and left view) respectively. For column interleaved displays, odd and even columns represent different views. Thus, only half of the vertical resolution per view is left for line interleaved display and half of the horizontal resolution per view is left for column interleaved display.

- **Active shutter technique**: is also called temporal multiplexing technique. The left and right images are displayed in the screen alternatively. Viewer need to wear a pair of glasses with active shutter. When the presentation of the left image, the active shutter dedicated to the right eye is close and vice versa. Compared with polarized technique, active shutter technique is less sensitive to head position change. It can provide two views with full spatial resolution in television application but only half temporal resolution (in case of frame-compatible format). Moreover, the luminance reduction and crosstalk may increase in comparison with polarized technique (Woods, 2001).

- **Eye wear and Head Mounted Display (HMD)**: this display technique is generally glasses composed of two miniaturized displays (LCD, OLED, etc.) and associated optic elements. They can provide full temporal and spatial resolution to each eye. However, it required precise calibrations (position, color, luminance and etc.) between the two displays. The other challenge is that the spatial resolution of the mini LCD sensor is not comparable to large LCD panels.

The main advantage of the visualization systems requiring glasses is the backward compatibility with conventional 2D content, i.e., take off the glasses, it turns back to a general 2D HD screen. The weaknesses are also apparent. The added optical instrument such as filters and glasses require precise synchronization and alignment, otherwise, view asymmetries problems will happen. Moreover, in television application, spatial resolution is reduced in the polarized case and temporal resolution is reduced in the active shutter case when using frame compatible formats. In all cases, participants watch the same couple of stereoscopic images even when moving in front of the screen. In that case, geometrical deformations of the visualized space differ from natural vision. It implies an artificial perception of depth variation and can lead to contradiction with other human senses like vestibular and hearing system (Shibata et al., 2011a, Shibata et al., 2011b).

The stereoscopic visualization techniques without glasses are also called autostereoscopic techniques. Each image is spatially oriented to the left or the right eye using appropriate filters. Thus, users don't have to wear specific glasses to watch 3D content. The main autostereoscopic techniques are:

- **Parallax barrier**: is composed of a layer of material with a series of precise slots, blocking light in certain direction using strips of black mask. The left and right views are represented by a different set of pixels. The parallax barrier radiates the set of pixels representing the left image to the direction only seen by the left eye and the set of pixels representing the right image to the direction only seen by the right eye.
- **Lenticular barrier**: uses cylindrical lenses instead of parallax barrier to radiate light in different set of pixels to different direction.

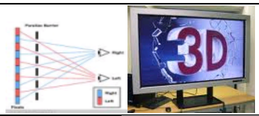
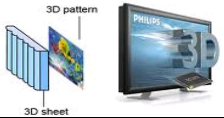
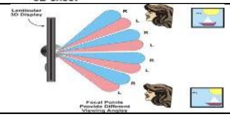
The drawback of this technique is that the viewing position is highly restricted. Viewer can only view the 3D correctly in a limited angle in front of the screen. In case of the wrong position or moving in front of the screen, the depth sensation can be lost or reversed (left image on the right eye and vice versa). Many autostereoscopic displays only support two views and only one user at a time.

Advanced autostereoscopic technique supports multi-views (more than 2). It allows multi viewers in the same time but in different position and it can also support motion-parallax features. However, the per-view resolution depends on the panel resolution and the amount of views. It can be very low based on current technique.

Moreover, eye tracking (Yong-Sheng Chen, 2001) or motion sensor (Jens Ogniewski, 2011) can be used to detect precisely the user position in front of the screen. View synthetic technique can re-render the left and right images corresponding to the current view position. These techniques can help to solve the positioning constraint problem. However, it increases the system complexity as well as requires reliable algorithm to render the synthetic views.

Table 1-3 summaries the principles, the advantages and the QoE issues of different stereoscopic 3D display techniques.

Table 1-3 : Characteristics of different display systems

3D display technique			Advantage	QoE issues
Glasses needed (stereoscopic)	Anaglyph		Compatibility with existing 2DTV and end-to-end architecture, easy to transmit and represent.	Color asymmetries, loss of luminance
	Polarized		Full temporal resolution, 2D and 3D smoothly exchange	Loss of luminance, crosstalk depending on filter adjustment, filter quality and viewing position
	Active shutter		Full spatial resolution, 2D and 3D smoothly exchange	Loss of temporal frequency, loss of luminance, alternate visualization between eyes, crosstalk
	Eye wear and HMD		Full spatial and temporal definition of each eye	Difficulties to calibrate two display, limited spatial definition of mini LCD sensor with current technique, unknown impact in human system (vestibular)
Glasses free (autostereoscopic)	Parallax barrier		No glasses needed.	Loss of luminance, loss of spatial definition of the panel, limiting sweet position of 3D viewing
	Lenticular sheet		No glasses needed.	The same drawbacks as the parallax barrier
	Multiview autostereoscopic		Support several viewers	The limited single view resolution with current technique and display resolution
	Autostereoscopic with eye tracking or motion sensor		No glasses needed, head movement(motion parallax) supported	High complexity to calculate the precise position and generate the synthetic views

1.6 Conclusion

In this chapter, we discussed the main challenges of the QoE assessment for S-3DTV. From foundation of depth perception and its potential impact on the QoE of S-3DTV, conclusions can be drawn as:

- People can extract the depth information from nine different information resources. Binocular disparity is one of them and it is particularly sensitive in personal space (less than 10 meters in depth). S-3DTV is able to provide the binocular disparity in the image. Thus, our vision system may take advantage of this additional depth cue to generate an enhanced depth illusion.
- Stereoscopic system is not a true representation of the real 3D world. The discrepancies between viewing S-3DTV and viewing real scenes have potential impact on the QoE. Visual discomfort or even visual fatigue will be induced if these discrepancies are larger enough.

New QoE assessment method should be able to highlight advantage as well as reveal problems of the QoE of S-3DTV.

From the review of QoE issues on S-3DTV broadcast chain, we conclude:

- There is no perfect solution proposed for S-3DTV broadcast which can provide a sufficiently high resolution to each eye without exhibiting view asymmetries. Various techniques are available and QoE related issues are not identical.

Thus, specification and measurement of these QoE issues are necessary in order to understand their quality impact, select the optimum solution and further improve the technique design.

Part I Towards methodologies for assessing S-3DTV QoE

Chapter 2 Methodologies for assessing 3D QoE

Contents

2.1 Introduction

As presented in Chapter 1, various QoE issues exist in different technique on S-3DTV broadcast chain. They have potential impacts on the final acceptance and success of S-3DTV services. Thus, evaluation of QoE of S-3DTV is urgent and important for many applications. For example, it can be used to ease the specification process for end-to-end application (e.g. determination of video bitrates, S-3DTV display techniques as well as video encoder tools and architectures).

In the scientific and industrial field, subjective assessment is the most direct way to evaluate the human QoE opinion. Conventional subjective assessment mainly focuses on the evaluation of picture quality. But concerning the QoE of S-3DTV, picture quality might not be a sufficient term to represent the QoE. For example, it cannot directly highlight the advantages such as enhanced depth perception and the problems such as visual discomfort for S-3DTV. Moreover, concerning the specification of subjective QoE assessment, existing assessment methods do not consider the new characteristic of S-3DTV.

In this chapter, first, we present the state-of-the-art of subjective QoE assessment for S-3DTV in Section 2.2. The conventional standardized ITU recommendations for evaluating the picture quality are presented as well as ongoing activities towards assessment of S-3DTV. Moreover, explorative studies in the literatures besides ITU studies for assessing the QoE of S-3DTV are presented and discussed. Second, we propose and discuss how to adapt the conventional quality assessment methods to evaluate the QoE of S-3DTV in Section 2.3. Our proposal mainly focuses on two parts: QoE indicators and common features of subjective assessment. For QoE indicators, evaluation of the QoE of S-3DTV requires to use multi-dimensional QoE indicators in order to highlight advantages and reveal problems. For common features of subjective assessment, the requirements of comprehensive adaption of conventional subjective QoE assessment method (ITU-R BT.500) for assessing the QoE of S-3DTV are proposed. Section 2.4 draws the final conclusion.

2.2 State-of-the-art: subjective QoE assessment for S-3DTV

2.2.1 ITU Recommendations

The standardization of subjective quality assessment methods has got a long history. Earlier in 1974, the International Telecommunication Union (ITU) published the recommendation ITU-R BT.500 – “Methodology for the subjective assessment of the quality of television pictures”. Until now, this recommendation has been revised several times. The latest version of this recommendation, ITU-R BT.500-11 (ITU, 2002), was published in 2002 and it is still the most famous and widely used recommendation in the field of image quality assessment. Moreover, in 2007, ITU published the ITU-R BT.1788 (ITU, 2007a) – “Methodology for the subjective assessment for video quality in multimedia application”. This recommendation describes non-interactive subjective methods for evaluating the video quality of multimedia and data broadcasting application comprising video, audio, still-picture, text and graphics. The main difference between the ITU-R BT.500 and ITU-R BT.1788 recommendations is the fact that BT.500 is focused on subjective assessment of television pictures, i.e., for large video format; instead, BT.1788 is focused on subjective assessment of video quality for multimedia, i.e., reduced picture format.

ITU-R BT.500 specifies the common features and assessment methods for subjective quality assessment as shown in Table 2-1. “Common features” are the specification of general conditions for subjective quality assessment. “Assessment method” is the protocol to evaluate the particular question for subjective quality assessment. ITU-R BT.1788 shares similar specification of BT.500 by adapting some features for multimedia application, e.g., viewing distance is more flexible as constrained (one to eight times of display height) or unconstrained (based on viewer’s preference).

Table 2-1 : Specification of subjective quality assessment in ITU-R BT.500

Common features	**General viewing condition**	Specification of environment luminance, display calibration, viewing distance and etc.
	Source signals	The reference signal should be of optimum quality of the television standard used.
	Selection of test materials	Particular kinds of test material should be used to address particular assessment problems.
	Range of conditions and anchoring	The test conditions should cover full range of scales or extreme examples should be used as anchoring conditions.
	Observers	Screening of viewer, expert or non-expert viewers, and the required amount of observers.
	Instruction for the assessment	Instruction of the question, the method, the grading scale, the sequence and timing.
	The test session	Duration of test session, random order if several sessions are necessary.
	Presentation of the results	The method to present and statistically analyze the results.
Assessment method	Particular method should be used to address particular assessment problems.	

2.2.1.1 ITU Common features

Various features in subjective assessment can affect the experimental results. Thus, ITU-R BT.500 specifies these common features as follows:

- **General viewing conditions:** different environments with different viewing conditions can affect the experimental results. ITU-R BT.500 specified the environment luminance (room lighting and chromaticity of background), screen luminance, display brightness and contrast calibration, display resolution review, viewing observation angle and viewing distance.
- **Source signals:** source signals provide the reference picture directly, and the input for the system under test. It should be of optimum quality for the television standard used. The absence of defects in the reference part of the presentation pair is crucial to obtain stable results.
- **Selection of test materials:** the number and type of test scenes are critical for the interpretation of the results of the subjective assessment. New systems frequently have an impact that depends heavily on the scene or sequence content. Thus, the number and type of test scenes should be selected so as to provide a reasonable generalization to normal programming. Measurement of spatial and temporal

perceptual characteristics of the scene can be used to indicate the complexity of a scene.

- **Range of conditions and anchoring:** because most of assessment methods are sensitive to variation in the range and distribution of conditions seen, judgment sessions should include the full range of the varying factors or extreme examples as anchors to cover the large quality range.
- **Observers:** at least 15 observers should participate. They should be non-expert. Prior to a session, they should be screened for visual acuity, color vision and other visual anomalies.
- **Instruction for the assessment:** assessors should be carefully introduced to the method of assessment, the types of impairment or quality factors likely to occur, the grading scale, timing. Training sequences demonstrating the range and the type of the impairment to be assessed should be used with scenes other than those used in the test, but of comparable sensitivity.
- **The test session:** a test session should last up to half an hour. “Dummy presentations” should be introduced to stabilize the observer’s opinion. If several sessions are necessary, a random order should be used for the presentations; but the test condition order should be arranged so that any effects on the grading of tiredness or adaption are balanced out from session to session.
- **Presentation of the results:** presentation of results must cover detail of the test configuration, detail of the test materials, type of picture source and display monitors, number and type of assessors, reference system used, the grand mean score for the experiment, original and adjusted mean scores and 95% confidence interval.

2.2.1.2 ITU Assessment methods

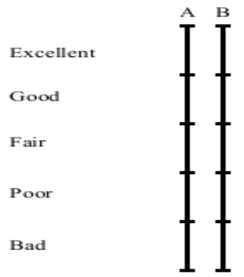
There are two classes of subjective assessment: (1) quality assessment is the assessment that establishes the performance of system under optimum conditions; (2) impairment assessment is assessment that establishes the ability of systems to retain quality under non-optimum conditions that relate to compression, transmission or etc.

ITU-R BT.500 also offers a collection of methods that are applicable for different assessment problems. In general, four different methods are proposed to assess the overall images quality of still images or short video sequences of 10 seconds: the double-stimulus-continuous-quality-scale method (DSCQS), double-stimulus impairment scales (DSIS), single-stimulus methods and stimulus-comparison methods. The recommended rating scales for the above four methods are shown in Table 2-2.

- **DSCQS and DSIS:** in DSCQS, observers assess the overall image quality for a series of image pairs. Each pair consists of an unimpaired (reference) and an impaired image (test) with a length of 10 seconds per image. These two images are presented one by one twice. In the second time of images presentation, observers are asked to rate the overall quality of each image. The presentation structure is illustrated in Figure 2-1. DSIS is similar to DSCQS but using impairment scales.

- **Single-stimulus method and stimulus-comparison methods:** in single-stimulus method, observers assess the overall image quality of each image in the stimulus set individually without a reference. In stimulus-comparison scaling, a series of image pairs, including all possible combination of two images in the stimulus set or just a selected sample of all possible image pairs, are presented to the observers. In this procedure, observers compare the two images for each image pair and assign their relationship by comparison scale as shown in Table 2-2.

Table 2-2 : ITU-R BT.500-10 recommendation rating scales (ITU, 2002)

DSCQS continuous quality scale	Comparison scale of stimulus-comparison	
	-3	Much worse
	-2	Worse
	-1	Slightly worse
	0	The same
	+1	Slightly better
	+2	Better
	+3	Much better

Single stimulus quality scale		DSIS and single stimulus impairment scale	
5	Excellent	5	Imperceptible
4	Good	4	Perceptible, but not annoying
3	Fair	3	Slightly annoying
2	Poor	2	Annoying
1	Bad	1	Very annoying

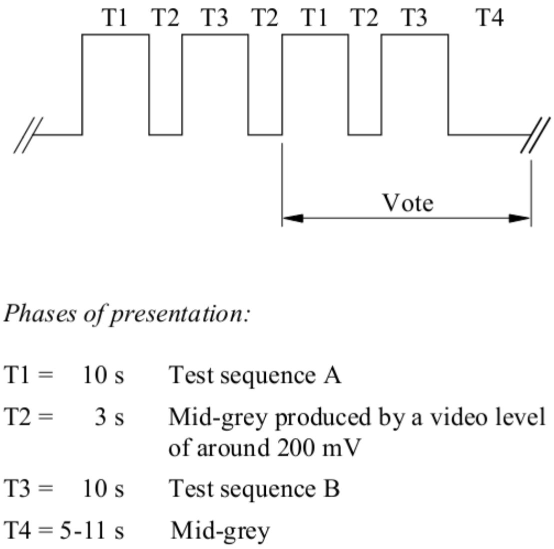


Figure 2-1 : Presentation structure of DSCQS and DSIS Variant II according to ITU-R BT.500-11 (ITU, 2002)

For assessment of longer video sequences (>60 seconds, up to 20 minutes), single stimulus continuous quality evaluation (SSCQE) and simultaneous double stimulus for continuous evaluation (SDSCE) methods are proposed.

- **SSCQE:** in SSCQE, observers continuously assess the picture quality of a long video sequence by moving a handset slider. The range (normally 0-100) of this slider is corresponding to the DSCQS continuous quality scales. SSCQE is used to assess video that contains scene-dependent and time-varying impairments.
- **SDSCE:** is similar to SSCQE, but with two stimuli presented at the same time. It is used to judge the fidelity between the reference video sequence and the test sequence. When the fidelity is perfect, the slider should be at top of the scale range (coded 100), when the fidelity is null, the slider should be at the bottom of the scale.

In ITU-R BT.1788, the Subjective Assessment Methodology for Video Quality (SAMVIQ) is proposed for the assessments of multimedia codecs or systems. It is derived from the DSCQS method of ITU-R BT.500. Blin in (Blin, 2006) stated that this method is efficient in the assessment of a large range of image quality as it provides reliable discrimination at both high and low quality levels.

- **SAMVIQ**: allows both hidden and explicit references in a multi stimulus test environment. Figure 2-2 illustrates a SAMVIQ test organization example. All the stimuli are directly accessible in a multi-stimulus form presenting multi buttons in Alphabetical order (i.e., A….Z). Besides the explicit reference, all the stimuli (Hidden reference and different algorithms) are assigned in a random order (represented by corresponding access buttons in Figure 2-2). The observer can choose the order of viewing the stimuli, review the stimuli as they want, and correct their votes, as appropriate. Each stimulus is thereby compared to an explicit reference which determines the best quality that can be achieved in the test. The observer rates using a slider on a continuous scale grading from 0 to 100 annotated by 5 quality items (Excellent, good, fair, poor, bad). For each stimulus, maximum length of 10 or 15 seconds is suggested as sufficient to get a stabilized and reliable quality score (Kozamernik et al., 2005, Blin, 2006). The quality evaluation is carried out scene after scene.

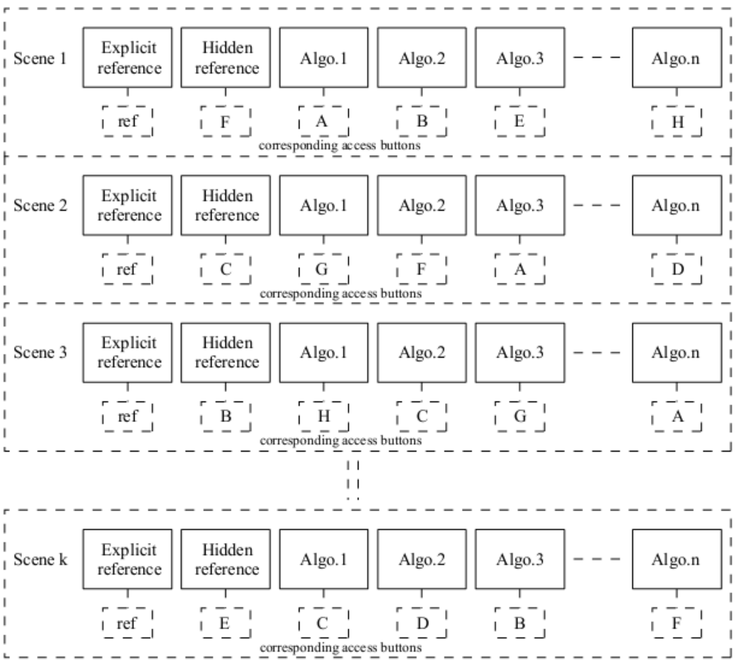


Figure 2-2 : A SAMVIQ test organization example (Blin, 2006)

2.2.1.3 ITU evolution towards assessment of S-3DTV

The original specification of ITU-R BT.500 does not cover the features of assessing S-3DTV. For assessing stereoscopic television pictures, ITU-R BT.1438 (ITU, 2000) – " Subjective assessment of stereoscopic television pictures" was published in 2000 by ITU. The main recommendations of ITU-R BT.1438 are:

- **Assessment factors:** besides the general factors applied to monoscopic television pictures (e.g., resolution, color rendition, motion portrayal, overall quality, sharpness), new factors peculiar to stereoscopic television system should be added, e.g., depth resolution, depth motion, puppet theatre effect, cardboard effect.
- **Assessment methods:** the methods of ITU-R BT.500-11 are also applicable in case of quality evaluation of stereoscopic images or videos.
- **Viewing conditions:** the display frame effect (i.e., windows violation in Chapter 1), inconsistency between accommodation and convergence (maximum value of depth of focus as ±0.3 diopters) and camera parameters (camera separation, camera convergence angle, focal length of lens) should be taken into account in determining viewing conditions.
- **Observers:** besides vision tests mentioned in ITU-R BT.500, stereopsis test should be used to screen the observers.
- **Test materials:** test materials for screening observers are recommended.

However, ITU-R BT.1438 still lacks specifications of many new characteristics of S-3DTV and how to assess them. Thus, ITU-R WP6 and ITU-T SG9 have addressed the requirements of developing a more adequate way to assess S-3DTV. The recent recommendations (draft) from ITU-R WP6 and ITU-T SG9 are listed in the table below:

Table 2-3 : Recommendation for subjective assessment of S-3DTV (NTT, 2011)

Recommendation	Title	Content
ITU-R BT.[3DTV SubMEth]	Subjective Methods for the Assessment of Stereoscopic Three-Dimensional Television (3DTV) systems	Recommendation covering subjective assessment methods for 3DTV
ITU-T P.3D-sam	Subjective assessment methods for 3D video quality	Recommendation regarding 3D assessment methods for the current 3D environment
ITU-T J.3D-fatigue	Assessment methods of visual fatigue and safety guideline for 3D video	Visual fatigue and safety assessment guideline for 3D video
ITU-T J.3D-disp-req	Display requirements for 3D video quality assessment	Requirements for displays used for 3D assessment testing

Meanwhile, Video Quality Expert Group (VQEG), the active contributor for most of the questions of ITU-T SG9, established a new project called "3DTV" targeting to investigate how to assess 3DTV subjective video quality.

2.2.2 Explorative studies

Besides the international standardization activities, in the last decade, many explorative studies towards better understanding and assessing the QoE of stereoscopic images have been done.

Pastoor in (Pastoor, 1992) discussed the human factors of 3DTV. He proposed that subjective evaluation criteria should be defined in order to guide the development of 3DTV services. Wöpking in (Wöpking, 1992) conducted a subjective experiment using single stimulus impairment scale to assess the annoyance of stereoscopic images with nine different disparity levels and five levels of resolution of background. IJsselsteijn et al. in (IJsselsteijn et al., 2000) investigated the effect of camera parameters and display duration on the subjective evaluation of stereoscopic images. They used single stimulus method with a numerical scale annotated from one to ten, where one represents the lowest level and ten represents the highest level of the scaled attribute. Observers were asked to rate quality of depth and naturalness of stereoscopic images. Yano et al. in (Yano et al., 2002) used the SSCQE method with a quality scale in their subjective test of visual comfort. Two 15 minutes video sequences, i.e., one 2D video and one stereoscopic video, were used as stimuli. Based on identifying underlying attributes of image quality and quantifying the perceived strengths of each attribute, Meester et al. in (Meesters et al., 2003a, Meesters et al., 2004) gave a discussion about how the principle of a quantitative quality measure of image quality for conventional 2D images can be applied in image quality research for 3DTV. Kooi in (Kooi and Toet, 2004) used DSIS Variant I method and an adapted five-level comfortable impairment scale (1 as Equal viewing comfort, 2 as Slightly reduced viewing comfort, 3 as Reduced viewing comfort, 4 as Considerably reduced viewing comfort, 5 as Extremely reduced viewing comfort) to assess the visual comfort induced by visual asymmetries of stereoscopic images. Yano et al. in (Yano et al., 2004) used a five-level visual fatigue scale (5 as I am not tired, 4 as I sense a little tired, 3 as I am a little tired, 2 as I am tired, 1 as I am very tired) and performed changes of accommodation and convergence to evaluate the view's subjective fatigue level after 1 hours of stereoscopic content viewing. Emoto et al. in (Emoto et al., 2004) proposed that the change of fusional amplitude and accommodation response is a valid indicator for visual fatigue. Seuntiëns et al. in (Seuntiens et al., 2005) used the single stimulus assessment method with a five-level quality scale to assess the naturalness and the viewing experience on 3D images. In order to investigate the perceptual attributes of crosstalk in 3D images, the same authors in (Seuntiëns et al., 2005) used the same single stimulus assessment method with five-level categorical scale to assess the perceived image distortion and the perceived visual strain. In (Seuntiëns et al., 2006), the authors still used single stimulus method but with different scales to assess the effect of symmetric and asymmetric JPEG coding and camera separation. The perceived overall image quality was rated in the ITU five-level quality scale and the experienced eye strain was rated in the ITU five-level impairment scale. The perceived sharpness and depth were rated using a numerical scale from 1 up to 5. No adjectives were used on the depth and sharpness scale. In his thesis (Seuntiëns, 2006), Seuntiëns summarized all his studies and tried to propose a perceptual model for 3D visual experience as shown in Figure 2-3.

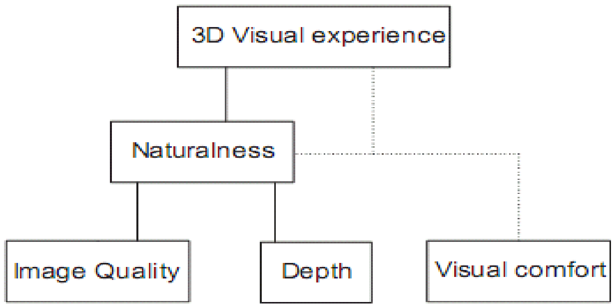


Figure 2-3 : Model of 3D visual experience (Seuntiëns et al., 2006)

In (Hyung-Chul et al., 2008), a questionnaire of five main factors of visual fatigue was proposed. In (Li et al., 2008), electroencephalography (EEG) signal was used to indicate visual fatigue. In (Lambooij et al., 2011), image quality, naturalness, depth percept and viewing experience of stereoscopic images with different camera baseline distance, different blur levels and different noise levels were rated using a single stimulus method with the ITU quality scale. Goldmann et al. in (Goldmann et al., 2010c, Goldmann et al., 2010a) established a stereo image and video database. They used a single stimulus method with a continuous quality scale to evaluate the quality of the stereoscopic images in the proposed database. Strohmeier et al. in (Strohmeier et al., 2010) used a mixed method approach combining psychoperceptual evaluation (Acceptance of quality, overall satisfaction, 3D impression) and qualitative attribute elicitation (perceived overall image quality and perceived depth) to get a more holistic understanding of 3D audiovisual quality of mobile 3D device. A “Paired comparison” method was used in (Barkowsky et al., 2009) to understand the influence of depth rendering on the quality of experience using an autostereoscopic display. In (Yamagishi et al., 2011), the authors assessed the perceived quality, depth and naturalness of the uncompressed and compressed stereoscopic images. They concluded that both perceived quality and depth are needed for assessing the 3D QoE. Naturalness was found to correlate highly with the quality.

The following table presents a summary of the studies presented above.

Table 2-4 : Overview of the explorative studies

QoE indicators	Methods	Scales	Studies
Texture Quality and Sharpness	Single Stimulus	ITU-R quality scale with or without adjectives	(Seuntiëns, 2006) (Seuntiëns et al., 2006) (Lambooij et al., 2011) (Yamagishi et al., 2011)
Amount of Depth	Single Stimulus	Numerical scale(0-5)	(Seuntiëns, 2006) (Lambooij et al., 2011) (Strohmeier et al., 2010)
Quality of Depth	Single Stimulus, Paired Comparison	Numerical scale(0-10)	(IJsselsteijn et al., 2000) (Barkowsky et al., 2009) (Yamagishi et al., 2011)
Visual comfort, Eye strain and Visual Annoyance	Single Stimulus, SSCQE, DSIS	ITU-R impairment and quality scale, adapted impairment scale from ITU-R	(Wöpking, 1992) (Yano et al., 2002) (Kooi and Toet, 2004) (Seuntiëns et al., 2006)
Visual Fatigue	Questionnaire, objective measurement (e.g., EEG)		(Yano et al., 2004, Hyung-Chul et al., 2008, Li et al., 2008, Emoto et al., 2004)
Viewing experience, overall image quality, visual experience	Single Stimulus	ITU-R quality scale	(Seuntiens et al., 2005) (Seuntiëns et al., 2006) (Seuntiëns, 2006) (Lambooij et al., 2011) (Goldmann et al., 2010c, Goldmann et al., 2010a) (Strohmeier et al., 2010)
Naturalness	Single Stimulus	Numerical scale(0-10), ITU-R quality scale	(IJsselsteijn et al., 2000) (Seuntiens et al., 2005) (Seuntiëns, 2006) (Lambooij et al., 2011) (Yamagishi et al., 2011)
Presence and enjoyment	Single Stimulus	ITU-R quality scale	(Seuntiëns, 2006)

2.2.3 Discussion

For ITU recommendations, the conventional standard like ITU-R BT.500 does not cover the new characteristics of S-3DTV. The adapted ITU-R BT.1438 only covers very limited new characteristics of S-3DTV. Thus, new questions for subjective assessment of S-3D video were addressed and new activities towards new subjective assessment methods for evaluating 3D video QoE are now making progress. The studies of this thesis also aim to contribute to the standardization of subjective QoE assessment methods for S-3DTV.

From the explorative studies towards assessing the QoE of S-3DTV, there are three main findings:

- Many studies used different QoE indicators, or subjective attributes (IJsselsteijn et al., 2002) to present the QoE of stereoscopic images including Amount of depth, Quality of depth, Texture Quality and Sharpness, Visual Comfort, Visual fatigue, Viewing experience (Overall Image Quality, or Visual Experience), Naturalness, Presence and Enjoyment. There are no common definitions for some QoE indicators. For example, Depth may refer to the amount of depth (Lambooij et al., 2011) or the quality of depth (IJsselsteijn et al., 2000). Image quality may refer to texture quality (Lambooij et al., 2011) or overall image quality (Goldmann et al., 2010c, Goldmann et al., 2010a). Thus, it may be difficult to make a fair comparison between studies. However, a common understanding towards assessing the QoE of S-3DTV can be drawn from explorative studies: conventional "quality" indicators are not enough to represent the QoE of S-3DTV. Consequently, multi-dimensional QoE indicators are required.
- The test environments among different subjective experiments were different. For example, concerning general viewing conditions, various types and size of S-3DTV display were used without any specification of the calibration process and the maximum luminance. The rule of determining the viewing distance varied from studies. There was also sometimes a lack of specification for test materials. Most of the studies did not follow the ITU-R BT.500 and ITU-R BT.1438 recommendations. This may be because the general viewing conditions proposed by ITU-R BT.500 are not adapted to 3D application. It may also induce difficulties for result comparison among studies.
- For visual fatigue measurement, there is still no common method to assess it.

The development of new standardized subjective QoE assessment method should consider the above three problems to provide specification to guide subjective assessment and to achieve reliable, comparable and repeatable subjective experiment results.

2.3 Towards comprehensive adaptation of subjective QoE assessment for S-3DTV

As discussed in the previous section, conventional subjective quality assessment methodologies need to be adapted to assess the QoE of S-3DTV. As QoE is multidimensional for S-3DTV, multi QoE indictors are required to represent the QoE of S-3DTV. Moreover, the specification of common features for assessing S-3DTV images is required to consider the new factors of S-3DTV since they might have potential impacts on the QoE.

2.3.1 Proposal of QoE indicators

The traditional concept to evaluate QoE, i.e., the assessment of the overall visual quality, is not enough to highlight the advantages and to reveal drawbacks of stereoscopic images, e.g., image quality is not sensitive to perceived depth and visual comfort problems. One of the common understandings from the literatures is that the QoE of S-3DTV should be multi-dimensional. By summarizing the literature proposals as presented in the previous section, we propose to use the below QoE indicators to assess the QoE of S-3DTV:

- **2D Image quality:** is the quality of rendering of textures and motions. In case of 2D image, 2D image quality is identical to traditional "Image quality". However, in case of 3D image, 2D image quality is focusing on the judgment of texture quality excluding the quality of depth.
- **Depth quantity:** is the amount of the perceived depth using the combination of monocular and binocular depth cues.
- **Visual comfort:** visual discomfort is related to multi-symptoms, e.g., eye strain, dry eyes and fusion difficulties. Variation of visual comfort can be also perceived as the sense of vision difficulties.
- **Depth rendering:** is the quality of the perceived depth, depending on the subject's preference on the basic criteria related to stretching or compression of the depth and the shape of the objects.
- **Naturalness:** focuses on the evaluation of the natural appearance of images, i.e., whether the scene is more or less representative of reality.
- **Visual experience:** is the overall quality of experience (QoE) of the images in terms of immersion and the overall perceived quality.

The above indicators aim to assess short term or instant opinion of QoE of stereoscopic images. Concerning the long term effect of viewing S-3DTV images, as presented in Chapter 1, visual fatigue might be induced and influence the QoE of S-3DTV. Thus, visual fatigue can be used as a long-term QoE indicator and defined as follows:

- **Visual fatigue:** is a decrease in performance of the visual system. It is an objectively and subjectively measurable criterion that is of particular value of ascertaining long-term adaptive processes of the visual system.

However, there are still not common agreements about how to measure visual fatigue. Chapter 4 in this thesis presents an experiment to investigate the visual fatigue of S-3DTV.

2.3.2 New factors affecting QoE assessment of S-3DTV

Concerning subjective quality assessment, common features as described in ITU-R BT.500 do not take into account new characteristics of S-3DTV. Thus, the adaption of conventional methodologies is required by considering new factors of S-3DTV. In this section, we discuss new factors affecting QoE assessment of S-3DTV based on the specification of ITU-R BT-500 recommendation as described in Table 2-1.

General viewing conditions

- **Luminance and contrast ratio**: additional optical instruments for 3D viewing, e.g., glasses and filters, cause a reduction of luminance. Our experiments (see Section 3.2) showed that up to 70% of luminance reduction occurs for active glasses 3DTV systems and about 50-60% was measured for polarized 3DTV systems. Thus, it seems mandatory that the peak luminance measurement should consider these aspects. In (Patterson, 2007), it is suggested that the minimum luminance for S-3DTV displays should be at least 30cd/m^2 to sustain depth of focus in order to guarantee basic depth sensation. Moreover, crosstalk is not only an annoying artifact but also influences the final contrast ratio. Thus, display measurement and calibration should be specified.

- **Background and room illumination:** if the position of the display is too close to the wall, objects with uncrossed disparity in the screen may appear to be inside the wall. This may cause conflict between the depth illusion from S-3DTV and reality. But some researchers also argued that it should not be a problem since people can recognize the S-3DTV display as a visual window. Further research is required to confirm this problem. Moreover, the room illumination may need to be defined more precisely regarding different 3DTV technique. For example, the lighting frequency of neon illumination depends on the local grid frequency. When using S-3DTV with active shutter solution, the interference between refreshment frequency of active shutter and the lighting frequency of neon illuminations source may induce serious flicking resulting in eye stress.
- **Monitor resolution:** overall display resolution, per view resolution, and stereoscopic resolution should be considered as aspects of the monitor resolution. Spatially multiplexed S-3DTV displays have reduced spatial resolution. Moreover, the physical pixel distribution may not be uniform or parallel. Time multiplex displays have reduced temporal resolution. Temporal asymmetries and temporal luminance distribution problems can also occur. It is still an open question how the viewer perceives these changes in resolution. The resolution in depth has been assessed in (Hodges and Davis, 1993), where the definition of perceived depth voxels and perceived depth range were introduced. In (Holliman, 2004a), stereoscopic resolution was defined as the number of planes of voxels within the certain depth range (±100mm around the display plane).
- **Viewing distance:** three times the height of the screen for HDTV and six times for SDTV were adopted as a recommendation in the ITU-R BT.710 (ITU, 1998) and ITU-R BT.500. However, manufacturers often recommend a designed viewing distance (DVD) which differs from the ITU standards. In some cases, e.g., autostereoscopic displays, 3D can only be watched at the DVD. Additionally, the Preferred Viewing Distance (PVD) was recommended in ITU-R BT.500 for 2D viewing in home environments. A subjective test had shown that PVD is a function of different parameters (Ardito et al., 1996) such as human visual acuity, screen size, picture resolution, etc. As explained in (Patterson, 2007), perceived binocular depth is a function of binocular disparity scaling by viewing distance and changing viewing distance will change the binocular depth perception. Thus, depth perception should be added as a new component for the PVD function.
- **Viewing position:** 3D geometrical distortions, e.g., shear distortion which is caused by a sideways movement of the observer (Woods et al., 1993), can influence the decision of viewing position. The reduction of luminance will become more severe when the observation angle increases. This also applies to motion parallax which is seen on multiview autostereoscopic displays. The viewing position is limited to certain positions in front of the display. If viewers are not in the right position, left and right view images will not be correctly perceived in the left and right eye. Crosstalk or reverses of left and right images may occur.

- **Depth rendering:** the way how a display represents the perceived depth based on the input video is defined as depth rendering. Depth rendering has been proved to significantly influence the quality of experience for autostereoscopic displays (Barkowsky et al., 2009). At the display side, depth rendering ability depends on the viewing distance, the content disparity, and the properties of the display (e.g., pixels sizes and allocation of pixels per view). Moreover, depth rendering should also consider constraints of the comfortable viewing zone. Further analysis about depth rendering ability in S-3DTV displays are presented and discussed in Section 3.3.

  - ***Source signals***

- **Video format:** various 3D representation formats are available in the literature such as conventional stereo video format, 2D-plus-depth-format, multi view video format (MVC) and multiview video plus depth format (MVD), layered depth video format (LDV) and depth-enhanced stereo (DES) format. For frame compatible formats such as Top-and-Bottom and Side-by-Side, the resolution reduction may affect the quality and require further investigation. For depth-map based formats, even for the LDV format, the quality of reconstructed novel view is still not comparable to native stereo views (Kauff et al., 2008, Kerbiriou et al., 2010). Specifications of video format and view synthesis algorithm are required.
- **Video format conversion:** the conversion between the aforementioned video formats is lossy in most cases. For example, a systematic loss of information for occluded objects occurs if 2D-plus-depth-format with a single layer of depth to is converted to conventional stereo video format (Kauff et al., 2008). Moreover, the amount of loss depends on the implementation used. A minimum accuracy for the format conversion should be defined, e.g., by providing a validation test set.

  - ***Selection of test materials***

- **Video content complexity:** for 2D video, the ITU-T P.910 (ITU, 1999) defines the spatial perceptual information (SI) and the temporal perceptual information (TI) as main elements of 2D video complexity. Some new measurements, e.g., called depth perceptual information (DI), should complement these two measurements. Regarding DI, spatial and temporal maximum disparity and average disparity in pixels may be considered. Adding a third dimension to the video content complexity also requires more standardized video sequences, e.g., further shooting sessions are required in order to generate the new reference scenes with various complexity levels considering SI, TI and DI.

  - ***Observers***

- **Number:** the number of observers depends upon the sensitivity and the required reliability of the experiments. As explained in (Ukai and Howarth, 2008), inter-individual differences in susceptibility when watching stereoscopic motion images are still unclear. The viewers' opinion was reported to be not as stable as in 2D. Thus, an increase of the number of observers might be needed to guarantee the reliability of the test, i.e., the minimum number of 15 observers recommended in ITU-BT.500 may not be sufficient.
- **Viewer’s stereopsis performance:** about 10-15% of the population cannot well perceive binocular depth cues, therefore additional optometric tests should be used to evaluate the viewer’s binocular vision performance. ITU-R BT.1438

recommends different vision tests (VTs) for assessing binocular vision performance of viewers.

- ***The test session***

- **Viewing duration:** 10s is used as a reference value in ITU-R BT.500 for short duration samples of 2D video. For the transition to 3D, there are two conflicting arguments. The first states that since S-3DTV is closer to the human natural viewing behavior, less time is needed to judge the quality. The second states that since viewer accustoms to watching 2D television, more time is needed since more information is contained in the additional dimension of S-3DTV. In (IJsselsteijn et al., 2000), for a short duration test, the presentation time had little effect on subjective evaluation results. However, in their experiment, only 5s and 10s were tested. Further studies are still required to investigate the viewing duration's impact in subjective test.

- ***Test results analysis***

- **Viewer factor:** the statistical analysis needs to be reviewed in order to learn about the rejection of an incoherent viewer. For S-3DTV, the subjective test results may be more sensitive to inter-individual differences or preferences. Therefore, the analysis of multimodal viewer distributions might be required.
- **Multi-dimension indicator analysis:** using multi-dimensional indicators for the evaluation of 3D images calls for new methods for summarization, statistical analysis, and careful interpretation of the results. It may also lead to new concepts of objective models for 3D video quality.

- ***Test methods***

- **Visual fatigue:** is an objectively measurable quantity. There are several measurement techniques proposed to assess visual fatigue, including optometric tests of the visual function, electroencephalography (EEG) and event-related potential (ERP) (Li et al., 2008), and eye tracking considering visual interest. These efforts may lead to standardized procedures and recommendations. An experiment combining objective and subjective measurement for long term 3D viewing is presented in Chapter 4.
- **Subjective QoE indicator:** multi-dimensional QoE indicators as proposed in Section 2.3.1 should be used to assess the QoE of S-3DTV. Particular indicators should be addressed to assess particular problems of S-3DTV. Moreover, interactions between different QoE indicators should be well specified.

The new factors affecting the subjective assessment for S-3DTV are summarized in Table 2-5. Most of the new factors require further experiments for specification purpose. Several studies in this thesis are targeting to contribute to this specification process to develop a new subjective QoE assessment method for S-3DTV.

Table 2-5: New factors affecting subjective assessment for S-3DTV

Feature	Factors	New factors
General Viewing Conditions	**Luminance and contrast ratio**	Luminance reduction caused by additional optical instrument, minimum luminance necessary to sustain DOF, contrast ratio affected by crosstalk
	Background and room illumination	Minimum distance between display and background necessary, technology of room illumination critical
	Monitor resolution	Recommendation of minimum values for spatial and temporal per view resolution and stereoscopic resolution
	Viewing distance	Designed viewing distance (DVD) fixed by display manufacturer and adding depth perception factor into preferred viewing distance (PVD)
	Viewing position	Avoidance of 3D geometrical distortion, luminance reduction, suboptimal viewing position for autostereoscopic displays
	Depth rendering	Upper bounds for Depth Of Focus and binocular disparity
Source signals	**Video format**	Requirements for depth representation formats
	Video format conversion	Specification of accuracy for conversion
Selection of test materials	**Video content complexity**	Measurement tools for depth complexity of content
Observers	**Number**	Re-evaluation necessary to guarantee stability and reliability of results
	Viewer's stereopsis performance	measurement of stereopsis, accuracy, ocular differences, etc.
The test session	**Viewing duration**	Re-evaluation of duration for presentation, voting, session length
Test Results analysis	**Viewer factors**	Rejection criteria, detection of bimodal distributions
	Multidimension indicators analysis	Statistical methods for analysis, e.g., relation, interaction and combination of subjectively measured QoE indicators
Test method	**Visual fatigue**	Objective measurement of visual fatigue
	Subjective QoE indicator	Multidimensional QoE indicators

2.4 Conclusion

In this Chapter, from the review of QoE assessment methodologies, several finding are noted:

- Conventional subjective quality assessment methods are not sufficient to evaluate the quality of stereoscopic images. ITU and VQEG are still working on the new subjective quality assessment methods for stereoscopic images.

- For explorative studies, various QoE indicators were used. However, there were no common definitions of these QoE indicators. Moreover, the viewing environment or conditions varied between studies. It is difficult to compare the results from different studies.

The main conclusions and contributions of this chapter are:

- To summarize multidimensional QoE indicators and their definitions, including 2D image quality, depth quantity, visual comfort, depth rendering, naturalness, visual experience and visual fatigue.
- A focused discussion towards comprehensive adaptations of subjective QoE assessment for S-3DTV was presented. New factors were proposed to be considered for developing new QoE assessment for S-3DTV. It may contribute to judge the importance of and help to define the new subjective QoE assessment methodologies of stereoscopic images for 3DTV.

Chapter 3 Characterizing S-3DTV displays

Contents

3.1 Introduction

So far, there is no transparent display without any QoE issues available as presented in Chapter 1. Thus, characterizing the S-3DTV display is essential for selecting the optimum display for the experiment or adapting the display performance to be optimal. The display performance should also be considered for analyzing the experimental results of QoE assessment. However, conventional subjective quality assessment methods such as ITU-R BT.500 lack specifications of S-3DTV display. In this chapter, we focus on two of the most important factors for characterizing the S-3DTV display: luminance rendering and depth rendering.

3.2 Luminance rendering

In ITU-R BT.500, the requirements for display conditions are summarized inside the general viewing condition. Four main points related to luminance rendering for laboratory environment and home environment are recommended as shown in Table 3-1.

Table 3-1 : Suggested monitor performance specifications in ITU-R BT.500

laboratory environment and home environment	
Ratio of luminance of inactive screen to peak luminance:	*≤0.02*
Ratio of the luminance of the screen, when displaying only black level in a completely dark room, to that corresponding to peak white	≈0.01
Display brightness and contrast	*set up via PLUGE(ITU-R BT.814 (ITU, 2007b))*
Peak luminance	*200 cd/m^2*

The principle behind this specification is that luminance rendering affect the human perception of image on television and the judgment of image quality (Pappas and

Safranek, 2000). Thus, in order to guarantee the reliability of the subjective test results and the repeatability of the test, the above specification is recommended.

In case of S-3DTV display, this specification may still be valid. However, new characteristics and possible QoE issues of S-3DTV display require to be added into the specification. For example, the luminance rendering of stereoscopic 3D display system may need to consider two different conditions: 1) Luminance without glasses in 2D mode, for watching 2D content in S-3DTV display (if compatible); 2) Luminance with glasses in 3D mode. Moreover, crosstalk, due to the imperfect filtering of left and right images, is one of the potential problem of visual discomfort (Kooi and Toet, 2004). It requires to be defined and be reviewed clearly.

In this section, we propose characterization of luminance rendering for S-3DTV display. A simple experiment to characterize different displays is presented in order to justify the importance of characterizing the luminance rendering of S-3DTV display.

3.2.1 New characteristics of luminance rendering of S-3DTV display

The luminance rendering of S-3DTV is proposed to cover two types of characteristics as follows:

- 2D characteristics: require only one view's measurement;
- 3D characteristics: require more than one view's measurement.

Moreover, it is also important to distinguish the 2D mode and the 3D mode in the S-3DTV display. Most of the current 3DTV display techniques are extended or advanced version of 2D image display with the functionality of separating and delivering different views to human's left and right eyes. The 2D mode and the 3D mode are defined as:

- 2D mode: most of the current stereoscopic displays are fully backward compatible to display directly the 2D image signals. In this case, it does not require the viewer to wear glasses. Measuring the 2D mode performance of S-3DTV is identical to traditional performance measurement of 2D display. Only 2D characteristic measurements need to be considered.
- 3D mode: in case of stereoscopic display, 3D mode requires the viewers to wear the dedicated 3D glasses (e.g., polarized glasses or active shutter glasses) to separate and watch the left and right images correctly. To measure the display performance of 3D mode, the effect of glasses filters (e.g., loss of luminance and crosstalk) should be taken in to consideration, i.e., the measurement should be behind the glass from an observer's point of view. It should cover both the 2D characteristic per view with the effect of filters and the 3D characteristic of combining more than one view.

In summary, the measurement of 2D mode in S-3DTV display should cover:

- 2D characteristic

The measurement of 3D mode in S-3DTV display should consider both:

- 2D characteristic per view
- 3D characteristic

2D characteristics of luminance rendering for S-3DTV display include:

- **Luminance transfer function (gamma function):** the maximum luminance can be measured by outputting a 100% white signal to the screen. The minimum luminance can be achieved by sending a black level luminance signal. The luminance transfer function of the screen should be equivalent to those of a reference CRT with the rendering intent expected of a TV system. It is recommended that a gamma nominal value of 2.2 be used. The gamma function is related to gamma encoding and gamma decoding of the image, which requires compensating for properties of human vision – to maximize the use of the bits or bandwidth relative to how humans perceive light and color (Rogowitz 1998). If the gamma function in the display side is not correct, the mismatching of gamma encoding and decoding will happen. This might result in visual artifacts such as blocking artifact and quantization of luminance.

  The additional filters such as polarized filters or active shutter filters in stereoscopic display can reduce luminance. Thus, it is mandatory to measure the luminance behind the 3D glasses on the 3D mode. Moreover, some immanent luminance reduction function (e.g., to reduce the power consumption or to reduce crosstalk level) in S-3DTV display can be also clarified by gamma function measurement.

- **Color gamut**: is the portion of the color space that can be represented, or reproduced in the 3DTV display. The intention is that colors within the relevant system gamut should be reproduced such that the human eye perceives them to be identical to the presentation on an ideal CRT monitor. Color gamut is commonly represented in the CIE 1931 chromaticity diagram (Broadbent, 2004).

  The additional filters in the 3D glasses or the S-3DTV display may also affect the color gamut.

- **Resolution:** is the number of distinct pixels in an image that can be displayed. It can be an ambiguous term especially as the displayed resolution is controlled by different factors in cathode ray tube (CRT) (e.g., spot size and focus), flat panel (e.g., physical pixel) or projection displays using fixed picture-element (pixel) arrays. To measure the resolution reproduction ability of the display, Fresnel zone plate can be used.

- **Temporal performance, response time**: is the amount of time that a pixel in a monitor takes to go from one value to another and back again. It is measured in milliseconds. Lower numbers means faster transitions and therefore fewer visible image artifacts. Raise time (black to white), fall time (white to black) and Gray level response time (gray to gray) can be used to represent the response time.

- **Uniformity:** is the measure of the luminance distribution on the display panel. The uneven distribution of the luminance level across the screen may also induce visible artifact and of course affect the QoE. For conventional 2D display, the tolerance level of uniformity defect for CRT and LCD is different, as 20% for CRT and only 5% for LCD proposed in EBU-TECH 3320 (EBU, 2010).

- **Viewing-angle dependency:** In many applications, where the monitor is being viewed by more than one viewer, or the viewer is allowed to move freely, accurate picture reproduction over a range of viewing angle is of vital importance. Since most of the optical instruments in the S-3DTV display are sensitive to angle change, incorrect viewing position may result in luminance changes, color changes and even crosstalk.

3D characteristics of luminance rendering for 3DTV display include:

- **Crosstalk:** refers to the incomplete isolation of the left and right image channels so that one leaks or bleeds into the other – like a double exposure. Subjectively it is called ghosting.

  It is a critical issue which reduces the QoE of S-3DTV displays. It may also cause visual discomfort problems. The crosstalk occurs in various stereoscopic display by a wide range of mechanisms, including: time-sequential on PDPs and CRTs (phosphor afterglow, shutter timing, shutter efficiency), MicroPolar LCDs (polarization quality, viewing angle), time-sequential on LCDs (pixel response rate, update method, shutter timing and efficiency), autostereoscopic (inter-zone crosstalk), polarized projection (quality of polarizers and screens), anaglyph (spectral quality of glasses and displays) (Andrew, 2010).

  In (Fournier and Thierry, 1994, Fournier, 1995b), 0.2 to 5% of crosstalk level was mentioned as a visibility threshold range. Kooi in (Kooi and Toet, 2004) proposed <5% in low disparity and 5% in high disparity. Seuntiëns proposed 2% of crosstalk as a limit for natural image (Seuntiëns et al., 2005). However, the perception of crosstalk depends also on the luminance, contrast and disparity (Seuntiëns et al., 2005).

- **Viewing position dependency:** the viewing position dependency problem can be critical depending on the S-3DTV display technique. For example, head rotation will cause the failure of filtering each view in linear polarized glasses resulting in serious crosstalk. In case of autostereoscopic displays, due to the fact that the views are projected in different directions in front of the display, the correct visualization strictly depends on the correct viewing position. Wrong viewing position may result in crosstalk or exchange of the left and right view.

Besides the above 2D and 3D characteristics, some further features in S-3DTV displays are required:

- **Image format:** It is also important to check if the display supports the dedicated input format in order to select the appropriate display for subjective test. In High Definition Interface (HDMI) 1.4 (HDMI, 2009), specification of the 3D formats of input signals are defined. Various 3D video formats, e.g. Frame packing, Field alternative, Line alternative, Side-by-Side (Full), L+Depth, L+Depth+Gfx+G-depth and Side-by-Side (half) may optionally be transmitted. Thus, it is necessary to check the 3D format compatibility of the S-3DTV display.

- **Image processing functionality inside the display:** For example, image scaling functions are normally used to upsample or downsample the image in order to fill the screen resolution. It should be done in such a way as to avoid the introduction of artifacts, such as excessive ringing, aliases or banding. Another important function is the de-interlacing function in the display. This is the process of converting interlaced video into a non-interlaced form. Moreover, 3D format inter-conversion is also an important process function for S-3DTV display. The performance of the above image process functionality depends on the internal implementation algorithm. It also has a potential impact on visual artifacts or image distortion. Thus, it is important to check this functionality of 3DTV display in order to select the optimal display or identify possible problems. In case of not using the internal image process functionality, the optimal post production algorithm (i.e. providing optimal quality) for image processing should be proposed.

3.2.2 Case study

In this section, an experiment for measurement of different S-3DTV display is presented.

The main goals of measuring the performance of different S-3DTV display are:

- To justify that the performance of different S-3DTV displays can vary. Thus, measurement or calibration of the performance of S-3DTV display for subjective quality assessment is mandatory.
- To select the best display in terms of display performance to use as a visualization platform for the studies in this thesis.

ELDIM devices (Boher et al., 2012) were used to measure the display's characteristics. ELDIM Muratest solution was used to measure the color and luminance performance, and ELDIM Optiscope was used to measure the display temporal characteristics such as response time. ELDIM FDLITE device and a Digital Video System (DVS) media server (which was a hard disk raid based digital video record and player capable of playing real-time HD videos up to 1080p 60Hz) were used to output the image and video test panels.

Both of the passive and active 3DTV solutions were tested. In Table 3-2, the results of selected measurement items of four different displays are presented.

Table 3-2 : Measurement cases of 3DTV

	Display 1 (LCD+passive)		Display 2 V1 (LCD+passive)		Display 2 V2(1) (LCD+passive)		Display 3 (PDP+active shutter)	
Measure items	2D	2D per view	2D	2D per view	2D	2D per view	2D	2D per view
Luminance (Max)	285 cd/m^2	130 cd/m^2	147 cd/m^2	60 cd/m^2	250 cd/m^2	100 cd/m^2	60(2) cd/m^2	17(2) cd/m^2
Luminance reduction	54%		59%		60%		72%	
Luminance transfer function (Gamma function)	2.22	2.22	1.41	1.41	2.19	2.18	2.2/ 1.2(2)	2.2/ 1.2(2)
Response time	Raise: 4.3 ms Fall: 8.2 ms		Raise:4.6 ms Fall:5.3 ms		Raise: 4.7ms Fall: 5.4 ms		Raise: 0.045ms Fall: 17.74ms	
3D characteristic								
Crosstalk(3)	<4%		<3%		<3%		<3%	

(1) Display 2 V1 and V2 are the same display reference but in different versions. V1 was published half-year earlier than V2.

(2) Due to the luminance adaptation function on PDP display, the gamma curve is close to 2.2 before gray level 160. However, it becomes a constant value after gray level 160. Thus, approximation estimation of gamma is 1.2. The measurement of maximum luminance is also affected the dynamic luminance adaptation function on PDP display. Here the values are measured by outputting full white image signals to the display.
(3) The crosstalk value presented here is measured by sending a full white panel to the left view and a full black panel to the right view, crosstalk = the luminance level in right view(through glasses) / the luminance level in the left view (through glasses) (Andrew, 2010).

There are several noticeable finding from this case study:

- As shown in Table 3-2, the measured luminance reduction for the LCD plus passive polarized solution is about 50-60%; however, around 70% luminance reduction occurs for the PDP plus active shutter solution. Only the “Display 1” and “Display 2 V2” are able to provide 200 cd/m^2 in 2D mode, which is the required peak luminance for subjective quality assessment in home environment in ITU-R BT.500. However, in 3D mode, none of the four displays can provide 200 cd/m^2 in a single view. The maximum luminance varies from 17 cd/m^2 for “Display 3” to 130 cd/m^2 as “Display 1” in 3D mode.
- Considering luminance transfer function, the measured gamma values of “Display 2 V1” and “Display 3” do not equal to the standardized value 2.2.
- The measurement of response time was based on direct measurement of raise time (black to white) and fall time (white to black) by the ELDIM device. However, this method may not be appropriate to identify the temporal performance of PDP displays since the principle of temporal refresh is different compared to LCD displays.
- Although the same 3% crosstalk level in Display 2 V2 and Display 3 were measured, viewer experienced more crosstalk in Display 2 than Display 3 due to the higher luminance in Display 2.

Based on the results of Luminance rendering measurement, “Display 1” and “Display 2 V2” were selected as the visualization platform for most studies in this thesis because they can provide:

- Gamma value for luminance transfer function close to the nominal value 2.2. It indicates that the reproduction of luminance level is correct.
- The luminance of 2D mode is more than 200 cd/m^2 and 3D mode is more than 100 cd/m^2 for both of the display.
- 8 to 12 ms white-black-white response time.
- Crosstalk level is lower than 5%.

3.3 Depth rendering

Compared with 2D displays, the key element of S-3DTV displays is the ability to render the binocular disparity information to the viewers in order to enhance the depth perception. The depth rendering of the S-3DTV displays is related to the viewing environment (e.g., viewing distance), the properties of S-3DTV displays (e.g., pixel size and display size) and constraints of the human visual system (e.g., depth of focus). How to represent the depth rendering ability of S-3DTV is still an open question. In this section, a theoretical model for analyzing the depth rendering of S-3DTV is presented. Combining physical parameters and perceptual constraints, the depth rendering ability as well as the angular depth plane interval are defined. Based on the proposed definition, the depth rendering abilities for different types of stereoscopic displays are analyzed and discussed.

3.3.1 Modeling depth rendering of S-3DTV

A schematic diagram of the simplified geometry of stereoscopic depth perception on planar S-3DTV displays is shown in Figure 3-1. It combines the physical parameters of the viewing environment with the constraints of DOF and binocular disparity.

The physical parameters presented in Figure 3-1 are:

- ***Inter-pupil baseline***: is the distance e between the eyes of the observer. An average of 65mm is used in our calculations.
- ***Viewing distance***: is the distance d between the observer and the display plane.
- ***Pixel***: is assumed to be an idealized square pixel grid in this study. The width of a pixel is denoted as p_w.
- ***Stereoscopic voxel*:** is defined in (Hodges and Davis, 1993) as the region of uncertainty for an object located in depth. The volume is formed by the intersection of the lines of sight from each eye.
- ***Depth plane*:** is parallel to the display surface. It connects the centers of the stereoscopic voxels with the same screen disparity. Its horizontal resolution in terms of pixels is reduced by 2 pixels for each step from the display plane due to the non-ability to render the corresponding border pixel in the other view.

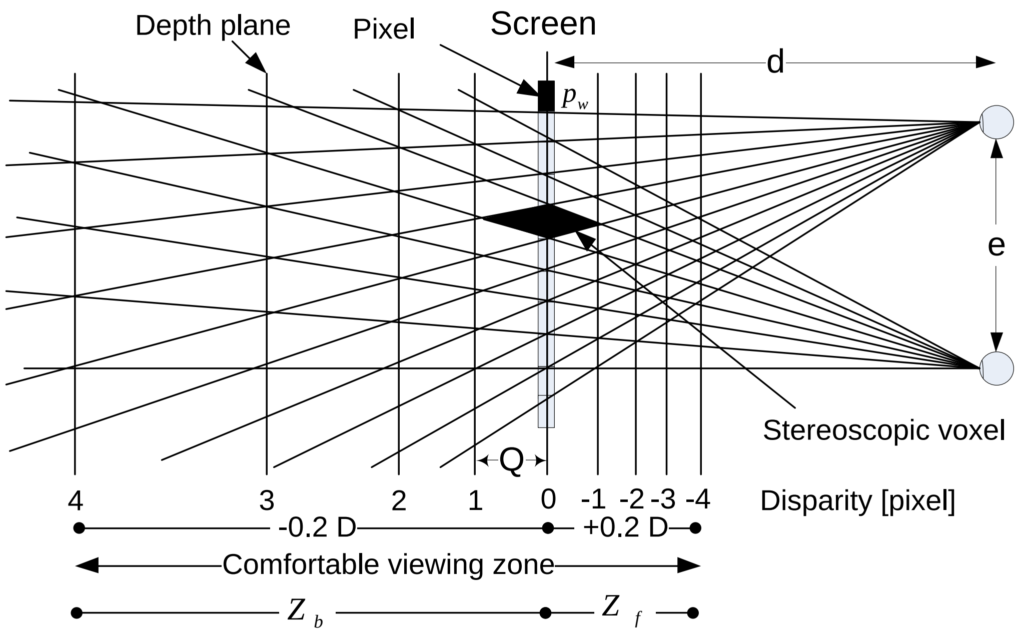


Figure 3-1 : Schematic diagram of physical and perceptual parameters of depth rendering (adapted from (Holliman, 2004a))

The perceptual constraints are:

- ***Depth of focus and Limit of Binocular disparity***: depth of focus refers to the range of distances in image space within which an image appears in sharp focus. It is usually given in terms of diopter. A value of ±0.2 diopters for the DOF was suggested in (Yano et al., 2004). The limit of binocular disparity is a region around the fixation point where disparities can still be comfortably fused. Its limitation is related to the human eye's aperture and depth of focus. Since the limit of DOF and binocular disparity resemble each other, they can serve as a general threshold.

- ***Comfortable viewing zone***: in (Lambooij et al., 2007), combining the limit of disparity and DOF, the authors determine a perceptual depth range where binocular fusion is possible and blur is not perceived so that stereoscopic visual comfort should be maintained. Calculated in distance, the comfortable viewing zone for disparity and DOF show very high resemblance and can serve as a general limit. Assuming the DOF equals to ±0.2 diopters, we can derive Z_f as the foreground distance of the comfortable viewing zone and Z_b as the background distance as a function of viewing distance d:

$$\boldsymbol{Z_f = d - \frac{1}{\frac{1}{d}+0.2} \quad Z_b = \begin{cases} \frac{1}{\frac{1}{d}-0.2} - d, \ if\ d < 5 \\ \infty, \ if\ d \geq 5 \end{cases}} \qquad \textbf{(3-1)}$$

- ***Max uncrossed disparity in pixels*** $\boldsymbol{D_b^{max}}$: divergence of the eyes beyond the infinite plane, e.g., beyond parallel view axis, is uncomfortable for the viewer. Thus, the maximum uncrossed disparity in pixels should be limited as inter-pupil baseline e divided by the width of a pixel p_w:

$$\boldsymbol{D_b^{max} = \frac{e}{p_w}} \qquad \textbf{(3-2)}$$

To combine the physical parameters and perceptual constraints, the below factors to represent the depth rendering ability of S-3DTV are defined:

- ***Depth rendering ability in pixels*** $\boldsymbol{D_f, D_b}$: is defined as the number of depth planes that can be represented within a comfortable viewing zone of display. $\boldsymbol{D_f}$ (the depth rendering ability in the foreground) and $\boldsymbol{D_b}$ (the depth rendering ability in the background) can be acquired as follows:

$$\boldsymbol{D_f = \frac{Z_f \cdot e}{(d - Z_f) \cdot p_w} \quad D_b = \frac{Z_b \cdot e}{(Z_b + d) \cdot p_w}} \qquad \textbf{(3-3)}$$

- Figure 3-1**Angular depth plane interval $\boldsymbol{Q_{angular}}$ (** as shown in Figure 3-1**):** is the distance between two adjacent depth planes, providing a measure of one step for depth quantization. The value stays almost constant if measured in angular units instead of meters. In (Pastoor, 1992) the authors suggested that less than 0.8 min of arc is needed in order to avoid a visible quantization in the depth rendering. $\boldsymbol{Q_{angular}}$ can be approximated as follows:

$$\boldsymbol{Q_{angular} = 2 \cdot [\tan^{-1}(\frac{e}{2d}) - \tan^{-1}(\frac{e-p_w}{2d})]} \qquad \textbf{(3-4)}$$

3.3.2 Analysis of depth rendering abilities of different S-3DTV displays

There are various types of stereoscopic displays based on different view separation techniques as presented in Table 1-3 (Chapter 1). However, from the physical point of view, the depth rendering ability and the angular depth plane interval are more related to the pixel size and organization of pixels in each pair of stereoscopic views. Thus, in this section, we categorize stereoscopic displays into four types based on their organization of pixels in each pair of stereoscopic views:

- **Full resolution display**: can deliver two full resolution images, one to each eye. Normally, these displays consist of two displays or one single display with temporal multiplexing. Examples are Desktop displays with the shutter glasses solution, two HD projectors for TV or home cinema, and two 2K projectors in cinema.
- **Line interleaved display**: spatially interleaves rows from the left and the right view. Thus, they only render half of the vertical resolution to each eye but they maintain the full horizontal resolution.
- **Column interleaved display:** spatially interleaves columns from left and right views and provides only half of the horizontal resolution.
- **Multiview autostereoscopic display:** contains more than two views and can support motion parallax. However, each view resolution generally equals to the full panel resolution divided by the number of views.

The characteristics of different displays regarding depth rendering abilities are given in Table 3-3. From Table 3-3, the depth rendering ability for different types of stereoscopic displays can be summarized as:

Full Resolution Displays

As shown in Table 3-3, Desktop and TV displays have around 80 depth planes within the visual comfort region, and their angular depth plane interval is close to the 0.8 min/arc. For digital cinema viewing conditions, the depth angular disparity per voxel is 3.3 which are likely to cause depth quantization artifacts. A resolution of at least 8192x4320 would be necessary to reach the limit of 0.8 arcmin in order to avoid discontinuous depth quantization.

Line Interleaved Displays

In terms of depth rendering ability and maximum disparity, it has a similar performance as the first two full resolution displays, since the binocular parallax only depends on the horizontal resolution. However, for each eye, half of the rows will be seen as dark stripes.

Column Interleaved Displays

Since the horizontal resolution is sub-sampled by a factor of two, its depth rendering ability is reduced. Moreover, it may have the same problem of visible dark stripes in the columns as described for the Line Interleaved Displays.

Multi-view Autostereoscopic Displays

Consequently, in case of a nine-view display, each view will only contain about 1/3 of the horizontal and 1/3 of the vertical resolution. The results show a medium level of depth rendering ability but only 21° for the field of view because the fixed viewing distance specification is five times the height. As the viewing distance increases, the range of visual comfort region increases as well. This partly counteracts the effect of sub-sampling in the horizontal direction. However, the field of view decreases leading to a lower sensation of presence.

Table 3-3 : Depth rendering abilities of different displays

Characteristic	**Full Resolution**			**Line Interleaved**(2)	**Column Interleaved**	**Autostereoscopic**
	Desktop	TV or home cinema	cinema	TV	Desktop	TV
Total resolution [pixel]	1680x1050	1920x1080	2048x1080	1920x1080	1280x1024	1920x1080
View resolution [pixel]	1680x1050	1920x1080	2048x1080	1920x540	640x1024	640x360
Display Height [m]	0.3	1.35	8	0.572	0.3	0.61
Pixel width [mm]	0.285	1.25	7.4	0.53	0.29	0.565
Viewing distance [m] (3)	0.9	4	8	1.6	0.8	3
Z_f / Z_b **[m]**	0.13(f)/0.19(b) (1)	1.78(f)/16(b)	4.9(f)/∞(b)	0.39(f)/0.75(b)	0.11(f)/0.15(b)	1.125(f)/4.5(b)
D_b^{max} **[pixel]**	227	52	8	122	110	38
$D_f + D_b$ **[pixel]**	40(f)+40(b)	41(f)+41(b)	14(f)+8(b)	39(f)+39(b)	17(f)+17(b)	23(f)+23(b)
$Q_{angular}$ **[arcmin]**	1.3	1.1	3.3	1.3	2.5	1.9
Field-of –view [degree]	35 °	35 °	87 °	37 °	39 °	21 °

(1) f for foreground and b for background
(2) Line interleaved display corresponds to Display 1 and Display 2 in Table 3-2.
(3) For viewing distance, the display specifications of designed viewing distance are followed if available. If no specification can be found, for HD resolution at home, it follows three times of screen height and for HD or 2K resolution at cinema, one times of screen height.

3.3.3 Discussion of the depth rendering of S-3DTV display

The depth rendering ability mainly depends on two parameters: the viewing distance and the properties of the display. It is apparent from Table 3-3, that the best solution in our comparison is the system based on the two HD projectors (Column 3 in able 3-3). It provides a reasonably good visual comfort region (1.78 meter in the foreground and 16 meter in the background) and enough depth planes (41 depth planes in the foreground and background, respectively). It also features a 35 ° field of view that is necessary to create a remarkable sensation of reality (Mitsuhashi and Yuyama, 1991). It can be considered as the reference system with optimal depth rendering ability.

For small size displays, e.g., the Desktop display with full resolution or TV with Line interleaved display as shown in Table 3-3, a larger viewing distance might have priority over the field of view in order to guarantee a wider comfortable viewing zone. Table 3-4 illustrates the depth rendering abilities of Desktop (Full resolution) (Column 2 in Table 3-3) and TV (Line interleaved) (Column 5 in Table 3-3) in case of viewing distance of 4.5 times of the display height. The comfortable viewing zone, depth rendering ability and angular depth plane interval are functions of viewing distance. Thus, increasing the viewing distance within an appropriate level can increase the range in depth of the comfortable viewing zone, allowing larger disparity level to be fused (e.g., for line interleaved TV, it increases from 41 depth planes to 61 depth planes in both foreground and background) and resulting in smaller angular depth interval (e.g., for line interleaved TV, it reduces from 1.3 arcmin to 0.82 arcmin). The main drawback of increasing viewing distance is the reduction of the field of view (e.g., for line interleaved TV, it reduces from 37 degree to 22 degree).

Table 3-4 : Depth rendering ability of Desktop (Full resolution) and TV (Line interleaved) in case of viewing distance as 4.5 times of display height

Display	Viewing distance [m]	Z_f / Z_b [m]	$D_f + D_b$ [pixel]	$Q_{angular}$ [arcmin]	Field-of – view [degree]
Desktop (Full resolution)	1.35	0.29/0.50	62/62	0.92	20 °
TV (Line interleaved)	2.57	0.87/2.71	61/61	0.82	22 °

Similarly, for multi-view displays, increasing the viewing distance will contribute not only to a comfortable viewing zone but also to a reduction of artifacts due to depth quantization.

Another possible problem is the mismatching between content disparity and depth rendering ability of S-3DTV. For stereoscopic production, often the left and the right view are recorded and stored in conventional stereoscopic format, e.g., frame compatible format as shown in Chapter 1. In this case, the content disparity range is fixed and cannot be modified without extensive and lossy processing. In Table 3-3, the depth rendering ability of each display provides as an upper bound of comfortable viewing for each display. When the disparity range of the content is outside the range indicated for each display, the observers might suffer from visual discomfort. On the opposite side, when the disparity range of the content is much smaller than the depth rendering ability, the viewers will perceive a poor depth effect. As the depth rendering ability spans a range from 22 pixels for electronic cinema to 82 pixels for the HDTV projector solution, it might be difficult to use the same content in a subjective experiment.

In terms of subjective video quality assessment, the selection of test materials should cover the principle that content disparity should be adapted to the depth rendering ability of the display. Moreover, analysis or comparison of subjective assessment results should also consider carefully these two factors. When the content disparity is higher than the depth rendering ability of the display, viewers may have difficulties to fuse the image.

3.4 Conclusion

In this chapter, we focused on the proposal for characterizing S-3DTV displays concerning luminance rendering and depth rendering.

For luminance rendering, new characteristics were presented and discussed. A case study comparing four different S-3DTV displays was performed in order to highlight the differences in luminance rendering between different S-3DTV displays. The results reveal that:

- The luminance reduction for LCD plus passive polarized solution is about 50-60%; however, around 70% luminance reduction occurs for PDP plus active shutter solution. Only two displays in this study can reach the maximum luminance as >100 cd/m^2.
- All the displays have different levels of crosstalk.

For depth rendering, we defined new factors to represent the ability of depth rendering for S-3DTV display by considering the physical parameters and the perceptual constrains. Based on the proposed factors and definitions, different S-3DTV displays were analyzed. The result analysis reveals that:

- Line interleaved displays have similar good performance in depth rendering ability as full resolution display. However, column interleaved displays reduce the depth rendering ability because their horizontal resolution are halved.
- Increasing viewing distance within an appropriate level can increase the range of the comfortable viewing zone in depth, allow larger disparity level to be fused and decrease the angular depth plane interval. The main drawback is the reduction of the field of view.

To summarize, characterizing the S-3DTV display is necessary since the luminance rendering and depth rendering performances of S-3DTV depend on various factors and can vary among different S-3DTV displays.

Chapter 4 Measurement of visual fatigue in optimal viewing condition of S-3DTV

Contents

4.1 Introduction

In Chapter 1, we presented the human factors related to the stereoscopic viewing in 3DTV. Especially in Section 1.4, focused discussions on the possible reasons of visual discomfort and visual fatigue as well as the related studies were presented.

Visual discomfort is more related to subjects' particular complaints caused by viewing S-3DTV image in short term. It can be caused by various reasons, for example excessive disparity, image asymmetries and a pulse motion in depth. Yano et al. in (Yano et al., 2002) used the SSCQE method with a quality scale in their subjective test of visual comfort. Two 15 minutes video sequences, i.e., one 2D video and one stereoscopic video, were used as stimuli.

Visual fatigue is a decrease in performance of the visual system. The accumulation of short-term visual discomfort may result in visual fatigue (Yano et al., 2002). Emoto et al. in (Emoto et al., 2004) proposed that the change of fusional amplitude and accommodation response is a valid indicator for visual fatigue. In (Hyung-Chul et al., 2008), a questionnaire of five main factors of visual fatigue was proposed. In (Li et al., 2008), electroencephalography (EEG) signal was used to indicate visual fatigue.

Even if excluding the effect of accumulation of short-term visual discomfort, there are still two different hypotheses for visual fatigue on stereoscopic viewing:

- The pessimistic hypothesis is that the current stereoscopic techniques were designed in the way that it originally disobeys the normal functionality of human system (Mikšícek, 2006). Thus, visual fatigue is an inherent and unavoidable problem for current stereoscopic techniques.

- The optimistic hypothesis is that the human visual system is at ease with adaptation and learning and can easily adapt to view behavior changes (Lambooij et al., 2009b). Under this hypothesis, if the stereoscopic images are presented within the acceptable range within which short term visual discomfort is not induced and accumulated, long term 3D stereoscopic viewing may just require a simple adaptation of the visual system which should not cause visual fatigue.

In order to judge the above hypotheses, an experiment of measurement of visual fatigue is presented in this chapter.

Compared with previous studies (Yano et al., 2004, Yano et al., 2002, Emoto et al., 2004, Hyung-Chul et al., 2008, Li et al., 2008), the novelties of this study are:

- Two one-hour sport contents, one in 2D and the other in 3D, are used as stimuli.
- The stimuli and viewing conditions in this study are selected to guarantee that the perceived depth is located within the comfortable viewing zone (±0.2 diopters). The motion in the stimuli is stable without a pulse movement in depth.
- The image asymmetries in the stimuli are corrected in post-production.
- Objective methods including vision test and Electroencephalography (EEG) measurement as well as subjective questionnaire are used to measure visual fatigue.

4.2 Objective and subjective methods

Three types of tests as illustrated in Figure 4-1 were designed in order to detect and measure visual fatigue objectively and subjectively: 1) A vision test, before and after the one hour content viewing session, 2) A questionnaire, before and after the one hour content viewing session, 3) A 16-channels continuous EEG signal measurement during the one hour video viewing session.

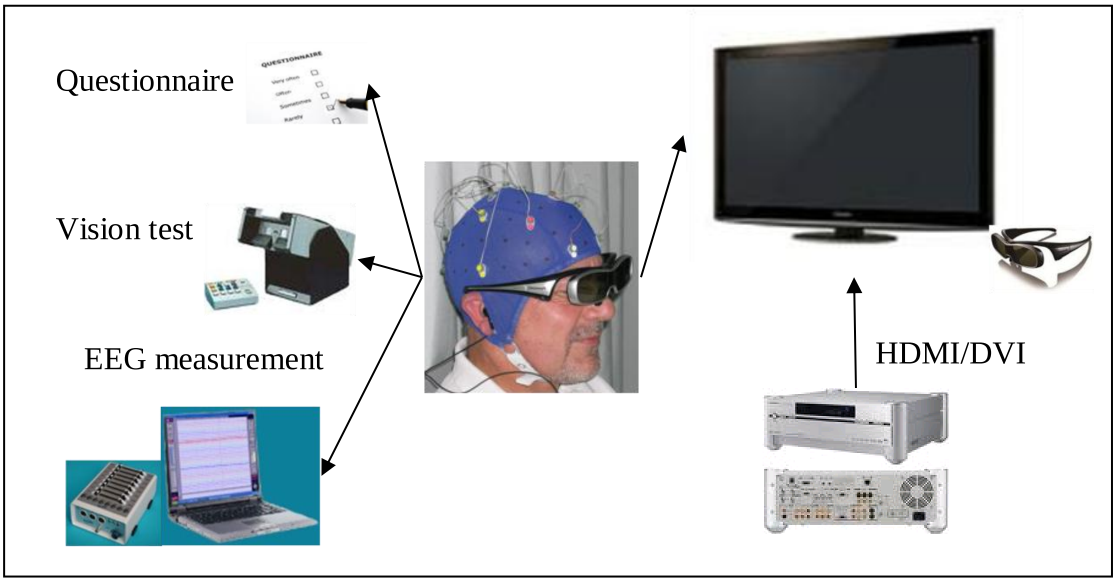


Figure 4-1 : Objective and subjective method for measure visual fatigue

4.2.1 Vision test

The vision test was implemented using Essilor ERGOVISION equipment. Six preset vision performance tests were selected as indicators of visual fatigue. The test principles are presented as follows:

- **Phoria** (both intermediate and far vision): is a latent deviation, or misalignment of eye that is only apparent some of the time. A phoria appears when fixation on a single object is broken and the eyes are no longer looking at the same object. It was used as an indicator of binocular vision problem and visual fatigue (Jiménez et al., 2000).
- **Fusion:** is the ability to fuse the image from the left and right eye to form a single vision. In (Lambooij et al., 2009a), the author reported that fusional amplitude is an efficient indicator for visual fatigue.
- **Monocular acuity** (visual acuity): is the visual acuity of left and right eye. In case of visual fatigue, the performance of visual acuity may be reduced.
- **Visual fatigue:** is a preset test in the ERGOVISION. By repeating the change from near vision to distant vision in a limited time (2s), the viewer is asked to report the presented 7 groups of numbers (each group is composed by five numbers). In case of visual fatigue, the viewers may not be able to react fast enough to report the correct number.
- **Stereoscopic acuity:** is the acuity to distinguish the disparity. 14 minutes, 7 minute, 6 minute, 3 minute and 1 minute of arc disparity level were measured.

4.2.2 Questionnaire

Kuze and Ukai in (Kuze and Ukai, 2008) developed a questionnaire to subjectively assess visual fatigue caused by viewing various types of motion images. Five factors including 1) Eye strain, (2) General discomfort, (3) Nausea, (4) Focusing difficulty and (5) Headache were reported as the effective indicators for visual fatigue. Visual function questionnaire (VFQ) with 25 items as proposed in (Mangione, 2000, Mangione et al., 2001), were used in (Lambooij et al., 2009a) to evaluate the visual fatigue for 3DTV viewing.

In our study, two questionnaires, i.e., a questionnaire before the one hour viewing session and a questionnaire after the one hour viewing session, were designed to evaluate the visual fatigue.

The questionnaire before motion image viewing covers five main parts:

- General health problems
- General Visual Health problems
- System Fatigue Symptoms
- Visual fatigue symptoms (direct): are more related to direct physical eye symptoms
- Visual fatigue symptoms (activity): are more related to ability and difficulties of completing certain vision task

The viewers were asked to fill this questionnaire before the test to report the health status, especially the visual health status before the test.

The questionnaire after the motion images viewing session consists of three parts:

- System fatigue symptoms
- Visual fatigue symptoms (direct)
- Visual fatigue symptoms (activity)

In this latter questionnaire, after the one hour long content viewing, viewers were required to report whether certain symptoms changed compared with the status before the motion image viewing session as well as the level of change in comparison scales with seven levels as illustrated in Table 4-1.

Table 4-1 : Comparison scales for visual fatigue symptom

Comparison scales for visual fatigue symptom	
Much more	7
Moderately more	6
Slightly More	5
Equivalent	4
Slightly less	3
Moderately less	2
Much less	1

Table 4-2 and Table 4-3 present the detail of the questionnaire before and after the one hour video viewing session respectively.

Table 4-2 : Questionnaire before one hour hours viewing session

Questionnaire before the test		
General Health problem		
1	In general, your health is	
	Excellent, good, fair, poor, bad	
General Visual Health problem		
2	Right now, your vision with both eyes is	
	Excellent, good, fair, poor, bad	
3	Do you worry about your eyesight?	
	No, slightly, moderately, much, hugely	
4	Does your vision prevent you from doing things?	
	Never, rarely, sometimes, Most of the time, always	
5	Do you have problems with near vision (reading, cooking, and sewing)?	Y/N
6	Do you have any problems with distant vision (TV, driving, sports)?	Y/N
7	Do you have vision problems in the dark?	Y/N
8	Do you have trouble driving at night?	Y/N
9	Do you have vision problems in sense of space (e.g., take something on a shelf)?	Y/N
10	Do you have trouble noticing objects to the sides when you walk?	Y/N
11	Do you have any difficulties, because of your view, to match your clothes?	Y/N
System Fatigue Symptom		
12	Do you have trouble in concentrating?	Y/N
13	Are you sleepy?	Y/N
14	Do you have a stiff neck?	Y/N
15	Do you have stiff shoulders?	Y/N
16	Do you have vertigo?	Y/N
17	Do you have nausea?	Y/N
18	Do you have a pain in the front of the head?	Y/N
19	Do you have a pain in the back of the head?	Y/N
20	Do you have a pain in the temples?	Y/N
Visual Fatigue Symptom (direct)		
21	Do you have heavy eyelids?	Y/N
22	Have you tired eyes?	Y/N
23	Do you have sore eyes?	Y/N
24	Is your vision obscured?	Y/N
25	Do you see blur?	Y/N
26	Do you see double?	Y/N
27	Do you have watery eyes?	Y/N
28	Do you have dry eyes?	Y/N
29	Do you have itchy eyes?	Y/N
30	Blink your eyes faster than usual?	Y/N
Visual Fatigue Symptom (activity)		
31	Do you have trouble focusing?	Y/N
32	Do you have the feeling that the movements of your eyes are decoupled?	Y/N
33	Do you have the feeling that your eyes look in different directions?	Y/N

Table 4-3 : Questionnaire after one hour video viewing session

Questionnaire after the test		
System Fatigue symptom		
1	Compared to before the test, you have (much more ...) issues of concentration?	Q*
2	Compared to before the test, you feel (much more ...) sleepy?	Q
3	Compared to before the test, you have (much more ...) stiff neck?	Q
4	Compared to before the test, you have (much more ...) stiff shoulders?	Q
5	Compared to before the test, you have (much more ...) dizziness?	Q
6	Compared to before the test, you have (much more ...) nausea?	Q
7	Compared to before the test, you have (much more ...) headache in the front of the head?	Q
8	Compared to before the test, you have (much more ...) headache in the back of the head?	Q
9	Compared to before the test, you have (much more ...) pain in the temples?	Q
Visual Fatigue Symptom (direct)		
10	Compared to before the test, you have (much more ...) heavy eyelids?	Q
11	Compared to before the test, you have (much more ...) tired eyes?	Q
12	Compared to before the test, you have (much more ...) eyes hurt?	Q
13	Compared to before the test, your vision is (much more ...) uncovered?	Q
14	Compared to before the test, you feel (much more ...) blur?	Q
15	Do you see double?	Y/N
16	Compared to before the test, you have watery eyes (much more ...)?	
17	Compared to before the test, you have eyes (much more ...) dry?	Q
18	Compared to before the test, you have (much more ...) itchy eyes?	Q
19	Compared to pre-test, do you blink faster than usual (much more ...)?	Q
Visual Fatigue Symptom (activity)		
20	Compared to before the test, you feel (much more ...) hard to focus?	Q
21	Have the feeling that the movements of your eyes are (much more ...) decoupled?	Q
22	Have the feeling that your eyes (much more ...) look in different directions?	Q
23	Is there a problem forced you to watch something other than the screen? Which one?	Y/N
24	Have you closed your eyes to re-obtain a clear vision for video?	Y/N
25	Have you found that the visual acuity test short / long-distance alternating fast (2 S) was (much more ...) difficult than before the video (2D/3D)?	Q
26	Have you found that the visual acuity test (reading the letters) was (much more ...) difficult than before the video (2D/3D)?	Q

*Q for comparison scales of seven levels as much less, moderately less, slightly less, No difference, slightly, moderately more, much more

4.2.3 EEG measurement

The cerebrum or cortex is the largest part of the human brain, associated with higher brain functions such as thought and action. The cerebral cortex is divided into four sections, called “lobes” as illustrated in Figure 4-2 (left):

- Frontal lobe: associates with reward, attention, short-term memory task, planning and motivation.
- Parietal Lobe: integrates sensory information from different modalities, particularly determine spatial sense and navigation.
- Occipital Lobe: is the visual processing center.
- Temporal Lobe: associates with perception and recognition of auditory stimuli, memory and speech.

The Brodmann area (Brodmann, 2006) as illustrated in Figure 4-2 (right) is a region of the cerebral cortex defined based on cytoarchitectonics, or structure and organization of cells. It is be widely used for approximating localization of brain activation.

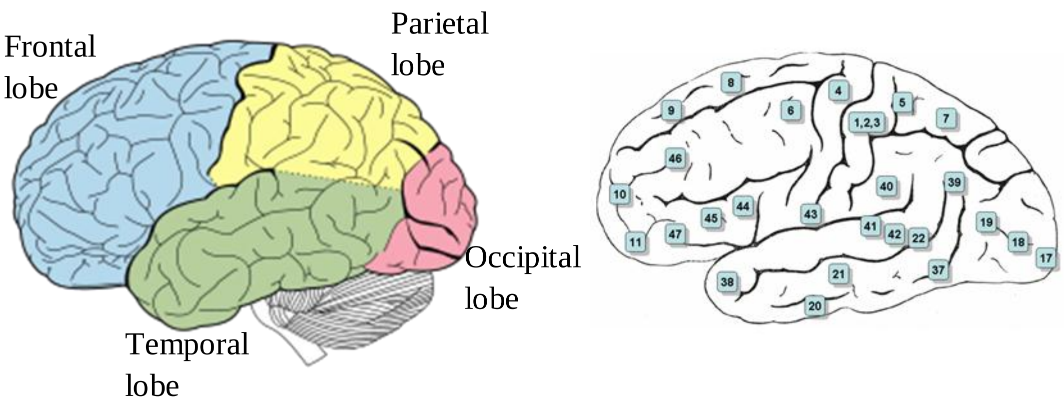


Figure 4-2: Principle Lobes of the cerebrum (left) and Brodmann area of lateral surface (right) (adapted from (Brodmann, 2006))

Brain activity measurement can provide information on changes in brain activity as a result of simultaneous behavior changes. EEG (Electroencephalography), which is the recording of electrical activity along the scalp produced by the firing of neurons within the brain, is widely used to represent the brain activity. The analysis of power spectrums of EEG signals frequencies is a common method to understand different levels of brain activity. There are five major brain waves distinguished by their different frequency ranges. These frequency bands from low to high frequencies respectively are called delta, theta, alpha, beta and gamma. Their characteristics (Sanei and Chambers, 2007) are illustrated in Table 4-4. For audiovisual activities, beta and gamma band would be the appropriate frequency band to be focused on. In (Li et al., 2008), the author reported that the power strength in beta band in most of the EEG channels increased as watching duration increased and it was much stronger in 3D than in 2D.

Table 4-4 : Characteristics of EEG frequency bands

Type	Frequency (Hz)	Activities
Delta	Up to 4	• deep sleep
Theta	4-8	• access to unconscious material • creative inspiration and deep meditation
Alpha	8-13	• relaxed awareness without any attention or concentration
Beta	13-32	• active thinking • active attention • focus on the outside world • solving concrete problems
Gamma	32-100	• occur during cross-modal sensory processing (perception that combines two different senses, such as sound and sight) (Sanei and Chambers, 2007) • short term memory matching of recognized objects, sounds, or tactile sensations

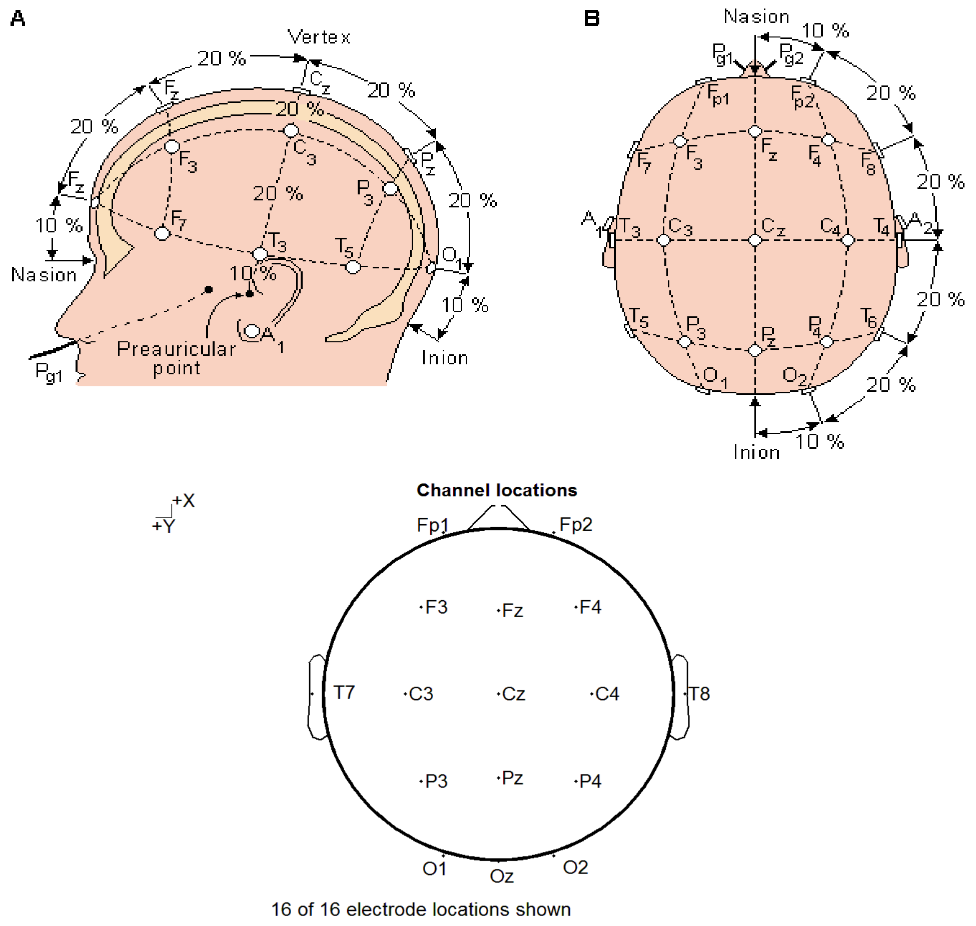


Figure 4-3 : The spatial location of EEG electrodes (top: international 10-20 system; bottom: 16 channel system in this study) (adapted from Fig. 13.2. (malmivuo and Plonsey, 1995))

In this study, a 16-channel Biosemi Active Two solution was used to record the viewer's brain activity during the one hour video viewing session. Electrode positions were a subset of the international 10-20 system sites as shown in Figure 4-3. The measured sixteen EEG channel included Fp1, Fp2, F4, Fz, F3, T7, C3, Cz, C4, T8, P4, Pz, P3, O1, Oz, O2, as F stands for Frontal, T strands for Temporal, P stands for Parietal, O stands for Occipital corresponding to the cerebrum lobes and C stands for central line. All the channel data were referenced to the Cz channel in the post-processing.

4.3 Experiment design

1) **Equipment:** The test was conducted in a test room, which is compliant with the recommendation for subjective evaluation of visual data issued by ITU-R BT.500. A 50 inch Panasonic 120HZ LCD display stereoscopic systems with active shutter glasses was used as the final visualization terminal (Display 3 in Table 3-2). This display was able to provide two full spatial resolutions to left and right eye with very low crosstalk level (less than 3%). An earplug was used to deliver audio signals. A digital video system (DVS) which can output 1920x1080 60HZ HD signal and stereo audio signals was used to deliver the uncompressed video content. All the 16 channels EEG signals were recorded by Biosemi ActiveTwo system in 512 Hz sampling rate.

2) **Observers:** 9 observers were recruited to participate in this test. All of them were non-experts in the audiovisual and video domain. All the viewers were healthy and without any system fatigue or visual fatigue symptom before the motion image viewing session.

3) **Stimuli:** The recorded Roland Garros Tennis Tournament videos were used as stimuli. One was the Men's Tennis Tournament Final captured in 2D condition and the other was the Women's Tennis Tournament Final captured in 3D condition. The reason to select different but similar content in 2D and 3D conditions was to avoid out of attention when viewing the same content twice. The content was captured carefully following the depth budget in order to render the perceived depth within the range of ±0.15 diopters (conservative value of the ±0.2 diopters). Thus, there were no excessive disparities. Moreover, the motion of the content was quite stable depending on the scene/camera. Shooting cameras were all fixed cameras and the zooming effects of cameras are rare in the shooting. Thus, no pulse motion in depth existed in the stimuli. Post processing was made by a professional company to get rid of the possible view asymmetries including geometrical asymmetry, luminance, and color asymmetry. In summary, the stimuli were selected in a way to avoid possible visual discomfort caused by content acquisition. The one-hour sequences with audio and video signals were stored and played in uncompressed HD format in order to avoid any compression distortion.

4) **Procedure:** For each subject, there were two sessions on different days in order to avoid the mutual interference between different session, one for the 2D condition and the other for the 3D condition. These two conditions were counterbalanced for each subject (e.g., 4 subjects did the 2D condition first and the other 5 did the 3D condition first). The subjects were not informed of the test conditions (2D or 3D) before the test. Moreover, the subjects were required to wear the glasses for both conditions (the glasses shutting function was forced to open in 2D).

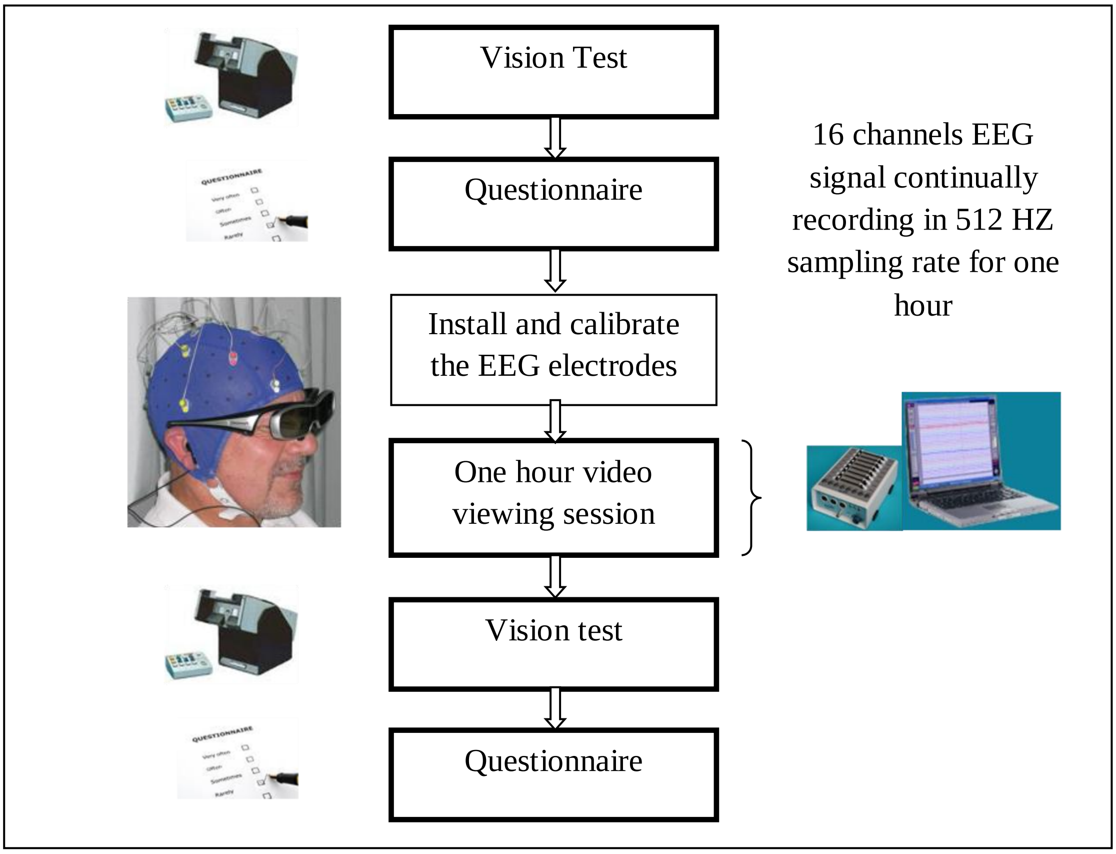


Figure 4-4 : The procedure of the experiment

For each session, the procedure is the same as illustrated in Figure 4-4 : firstly, the subject was required to take a vision test as presented in Section 4.2.1; second, the subject answered the questionnaire as presented in Table 4-2; third, the EEG electrodes were installed and calibrated following the manual of Biosemi Active Two system; fourth, the subject watched the video for one hour while the EEG signal was recorded continuously during the whole session; fifth, when the one hour video viewing session was finished, the same vision test was repeated again for the subject; sixth, the subject was required to finish the questionnaire as shown in Table 4-3. Moreover, the subjects were asked to report any subsequent symptoms that they experienced after having finished the experiments.

4.4 Result analysis

Concerning the vision test, the result analysis focused on whether the vision performance changes after motion image viewing session and whether this change are different comparing the 2D and 3D conditions.

The questionnaire before the test is carried out in order to reject the people who have system fatigue symptom or visual fatigue symptom before the test. Since the entire nine viewers reported neither system fatigue symptom nor visual fatigue symptom in the questionnaire before the test, no one was rejected. The questionnaire after the motion image viewing session directly reflected the change and the level of change of visual fatigue and system fatigue symptom. Statistical analysis was performed to find out whether there were significant changes on each visual fatigue symptom and whether there were individual differences among viewers.

For EEG signal results, first, a de-noising process as illustrated in Figure 4-5 was done to achieve a clean EEG data.

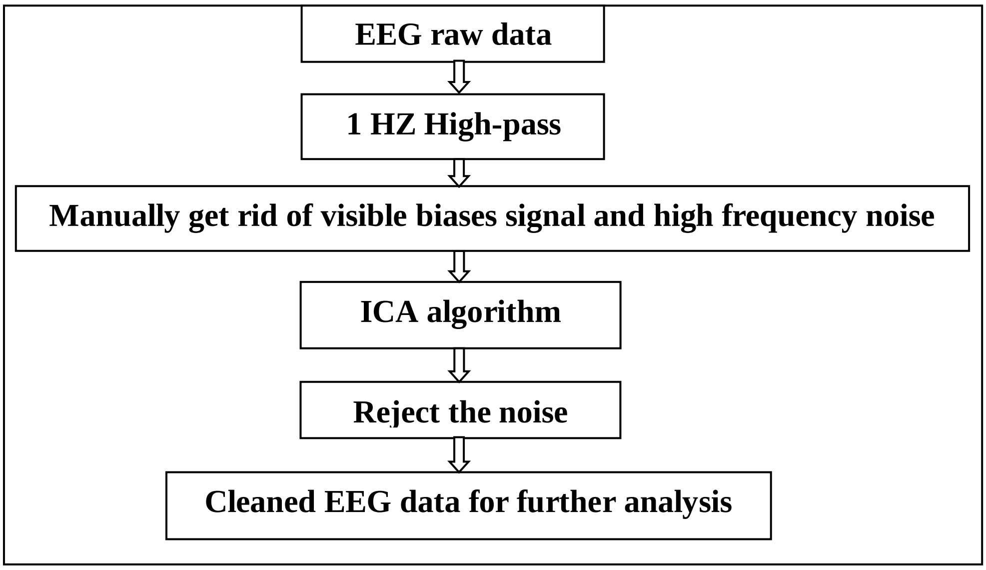


Figure 4-5 : De-nosing process for EEG data

EEGLAB (Delorme and Makeig, 2004), an open source toolbox for analysis of single-trail EEG dynamics, was used to implement all the de-nosing as well as the following statistical signal analysis of EEG data. For more details, the de-noising process included:

- 1 HZ high-pass filter: is used to remove the very low frequency signal which is mostly recognized as body movement or imperfect contact of the electrodes.
- Manually get rid of visible artifacts from EEG signal frames: this step is to get rid of the EEG signal frames which contain extreme values, abnormal trends or importable data by visual inspection. A simple example is illustrated in Figure 4-6. For the whole one hour's data, less than 1% of EEG data frames are removed in this step.

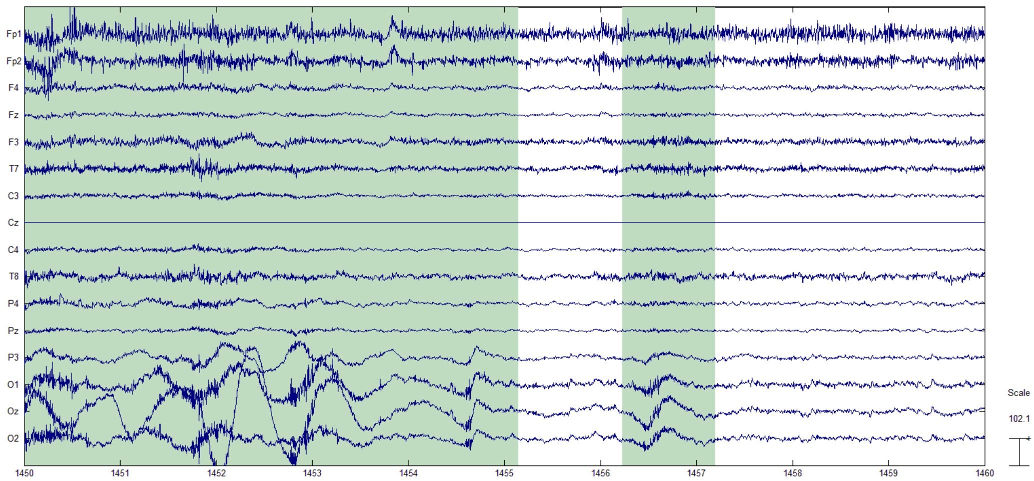


Figure 4-6 : Examples for visible artifacts (The marked green/grey parts of the EEG data frames are suspected to contain extreme values and abnormal trends)

- Independent component analysis (ICA): is a widely use method to decompose the times series EEG data into spatially stable mixtures of the activities of temporally independent cerebral and artifactual sources (Rogowitz 1998). EEG data can be roughly separated into three parts: independent components (ICs) accounting for brain and non-brain (artifact) processes, respectively, and smaller ICs whose maps and activities appear noisy and are poorly if all replicated from session to session. Ideally, only the brain ICs should be kept and all other components should be removed. It is still a widely open challenge to identify the artifacts for EEG processing.

- Reject the noise component: is to mainly to remove the common four types of non-brain ICs as illustrated in Figure 4-7, including eye blinks, lateral eye movement, electromyography (EMG) activity, and electrocardiographic activities. Other possible artifact ICs were judged and rejected by three rules: 1) Brain activity component should have a clear rhythm in 10 to 20 Hz while artifact component have sharp changes and huge variations on voltage; 2) Brain activity component should have a Gaussian distribution while artifact component normally have non-Gaussian distribution. One example is given in Figure 4-8.

The above de-nosing process was implemented to each one hour continuous recorded EEG data set. Thus, 9 viewers × 2 conditions, 18 data sets were processed. For each one hour data set, a maximum of five components are rejected.

The following sub section focuses on further analyzing and presents the results from different methods. A discussion of results is given at the end.

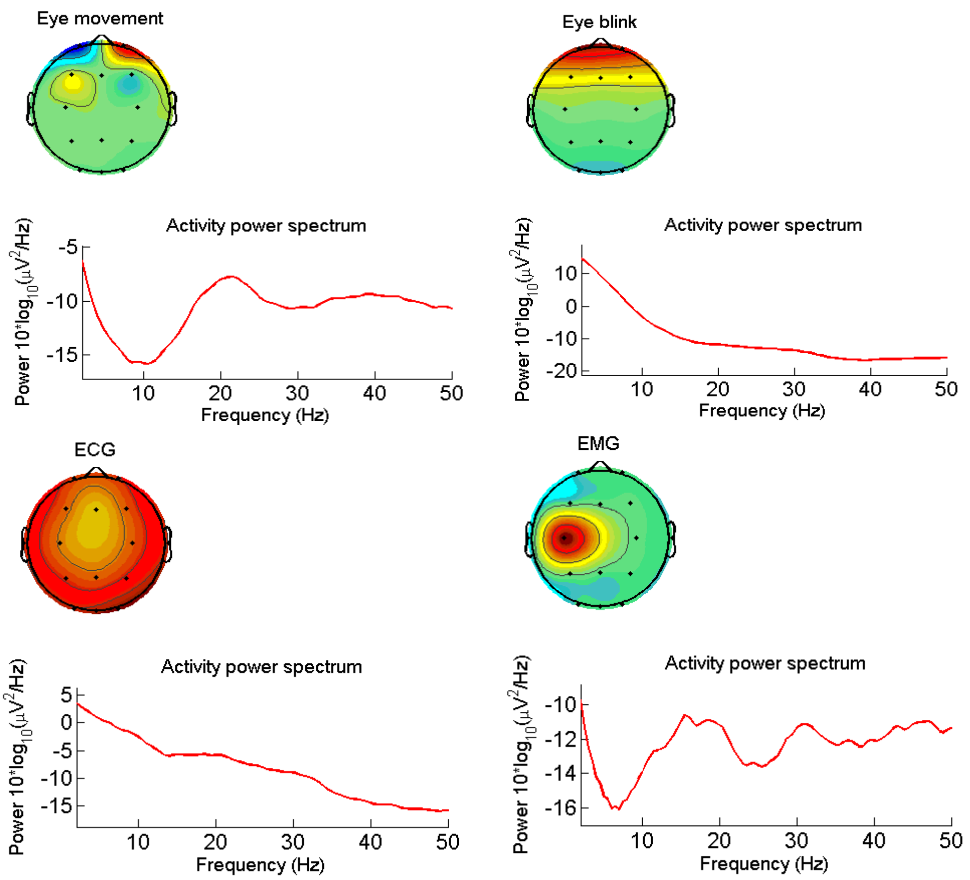


Figure 4-7 : Typical component properties of four non-brain ICs.

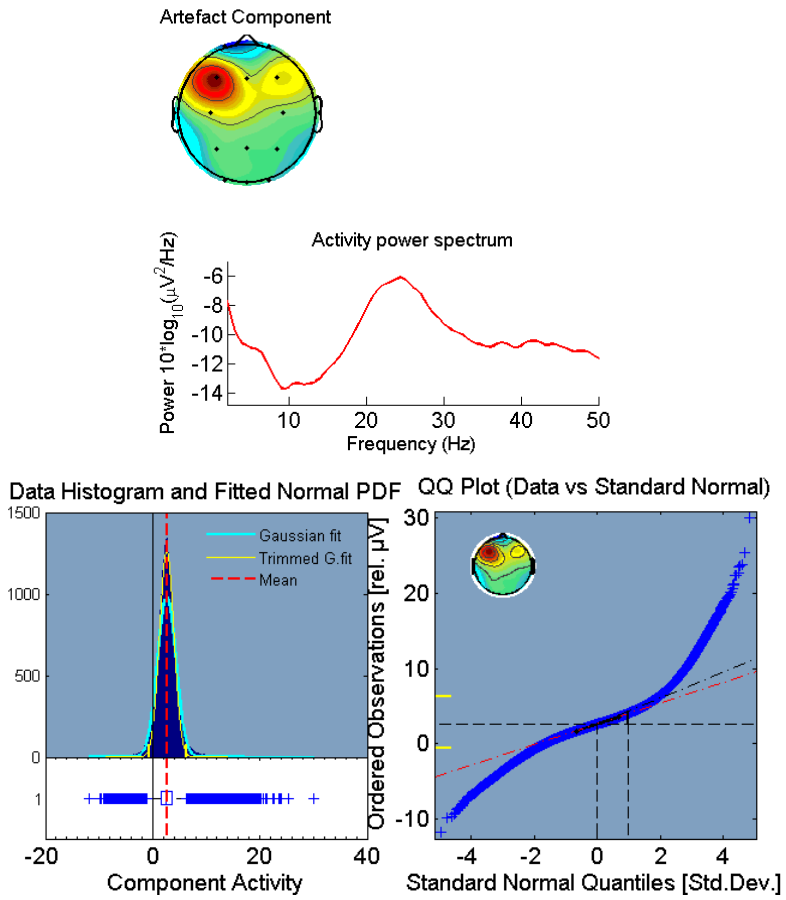


Figure 4-8 : Artifact component and its statistical analysis

4.4.1 Vision test

For the result analysis of vision test in 2D and 3D conditions, we are interested in two questions: 1) does the one hour video viewing session cause any performance change on the vision test; 2) comparing 2D and 3D conditions, are these changes different?

Thus, for 2D and 3D conditions, the performance change in each test item between the before and after vision test are illustrated in Table 4-5. “S” for “same” indicates that same performance is achieved; “B” for “better” indicates that better performance is achieved; “W” for “worse” indicates that worse vision performance is achieved.

From Table 4-5, we can observe that four subjects (Viewers 1, 2, 3 and 7) does not have any performance changes after the one hour video viewing session, independent of 2D or 3D condition. Viewers 4, 5 and 7 achieve better performance after the one hour video viewing session in some test items in both 2D and 3D conditions. This may be explained by training effect (Lambooij et al., 2009a). Only Viewer 8 and Viewer 9 had worse performance in Phoria (intermediate vision) test and stereoscopic acuity test, respectively. Paired Student T test were performed to compare the results of vision test (Assigning ‘S’ as 0, ‘B’ as 1 and ‘W’ as -1) shown in Table 4-5 between 2D and 3D condition for each test items. However, no significance differences (for all test items, p>0.17) were found.

Table 4-5 : Vision test results (Performance change between the before and after vision test) in 2D and 3D for 9 viewers.

Viewer / Condition		Phoria 1*	Phoria 2*	Fusion	Visual acuity	Visual fatigue	Stereoscopic acuity
Viewer 1	2D	S*	S	S	S	S	S
	3D	S	S	S	S	S	S
Viewer 2	2D	S	S	S	S	S	S
	3D	S	S	S	S	S	S
Viewer 3	2D	S	S	S	S	S	S
	3D	S	S	S	S	S	S
Viewer 4	2D	S	S	S	S	S	S
	3D	S	S	S	B	S	S
Viewer 5	2D	S	B*	S	S	S	B
	3D	B	S	S	S	S	S
Viewer 6	2D	S	S	S	S	S	S
	3D	S	S	S	S	S	S
Viewer 7	2D	B	S	S	S	S	S
	3D	S	S	S	S	S	S
Viewer 8	2D	S	S	S	S	S	S
	3D	W*	S	S	S	S	S
Viewer 9	2D	S	S	S	B	S	S
	3D	S	S	S	S	S	W

* Phoria 1 is Phoria test for intermediate vision and Phoria 2 is Phoria test for far vision.

*S for same, B for better, W for worse.

4.4.2 Questionnaire

For the questionnaire after the video viewing session, the comparison scale of 7 levels was normalized to value 1 to 7 as shown in Table 4-1. For each symptom, the mean opinion score and its confidence interval were calculated. Figure 4-9, Figure 4-10 and Figure 4-11 present the results of the questionnaire after the viewing session by simplifying the question into different symptoms. An ANOVA analysis was performed towards the question whether the difference between 2D and 3D conditions is significant for each symptom. No significant differences for any symptom with respect to the viewing condition (2D and 3D) were found ($p<0.05$ for rejecting the null hypothesis). However, three symptoms including the pain in the front head ($p=0.06$, $f=4$), the heavy eyelid ($p=0.11$, $f=2.8$) and the fatigue test difficulties ($p=0.11$, $f=2.8$) seemed to be more sensitive than the others. For each viewer, another ANOVA analysis was performed to compare the results between 2D and 3D viewing conditions. However, there were no significant differences reported. Generally, there were no significant evidences in the questionnaire showing that 3D viewing condition produces more visual fatigue compared to 2D viewing condition. Discussions with each viewer also confirmed these findings.

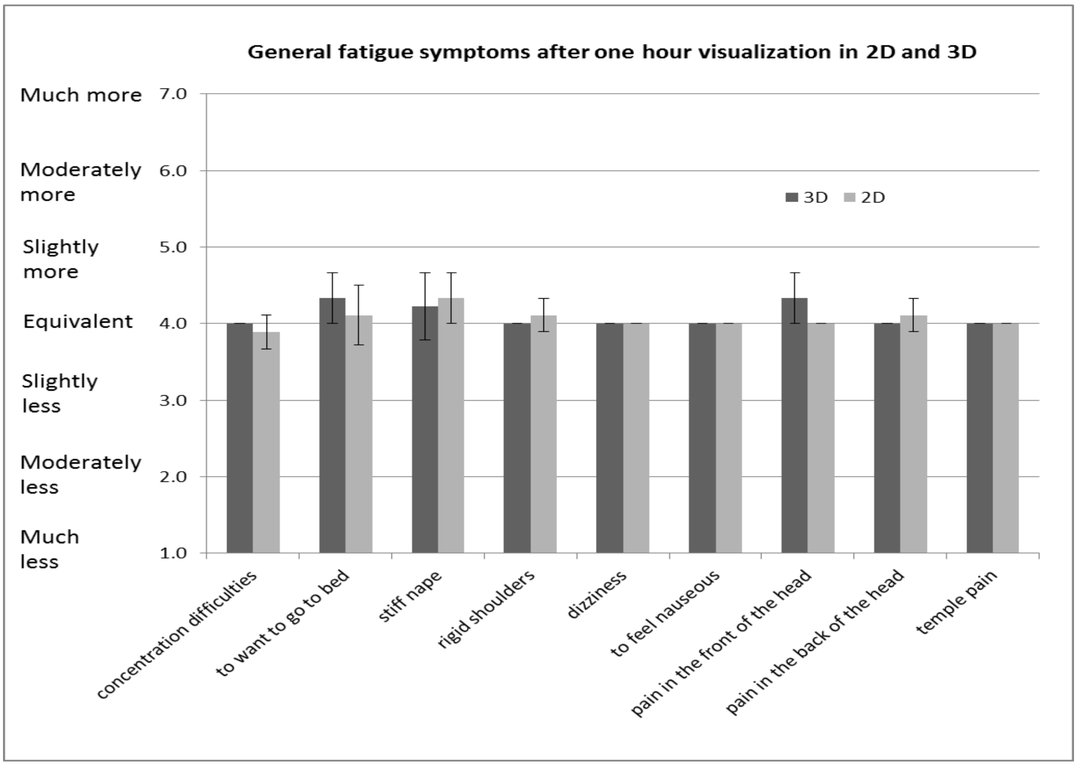


Figure 4-9 : General fatigue symptoms after one hour visualization in 2D and 3D

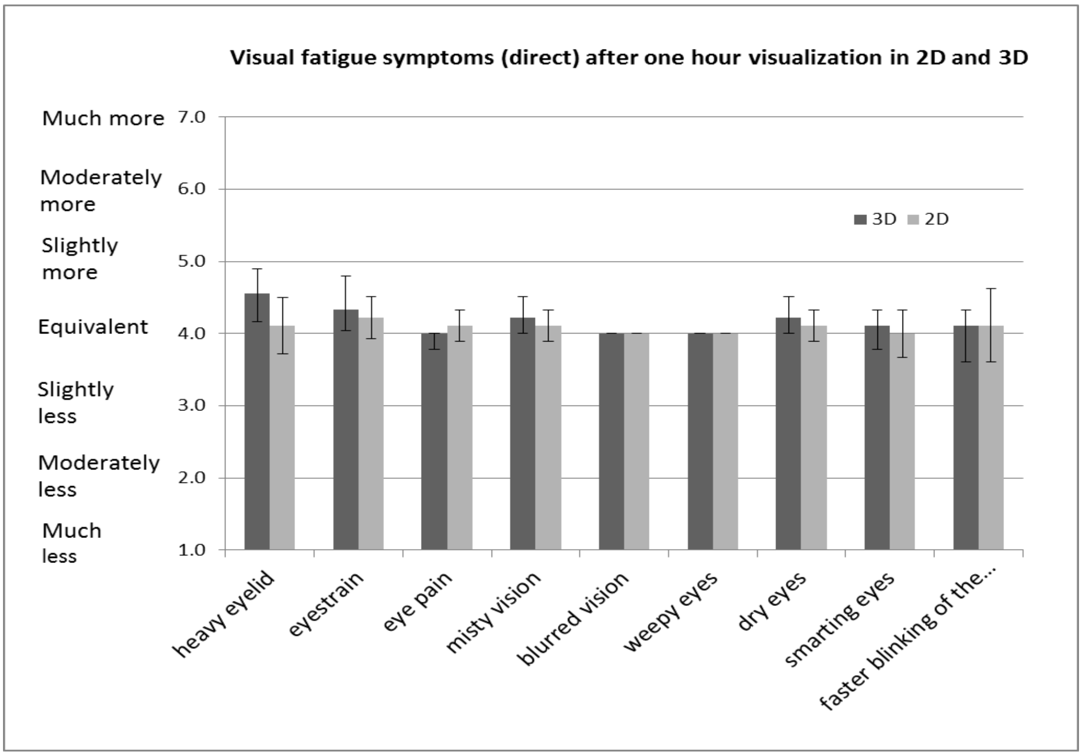


Figure 4-10 : Visual fatigue symptoms (direct) after one hour visualization in 2D and 3D.

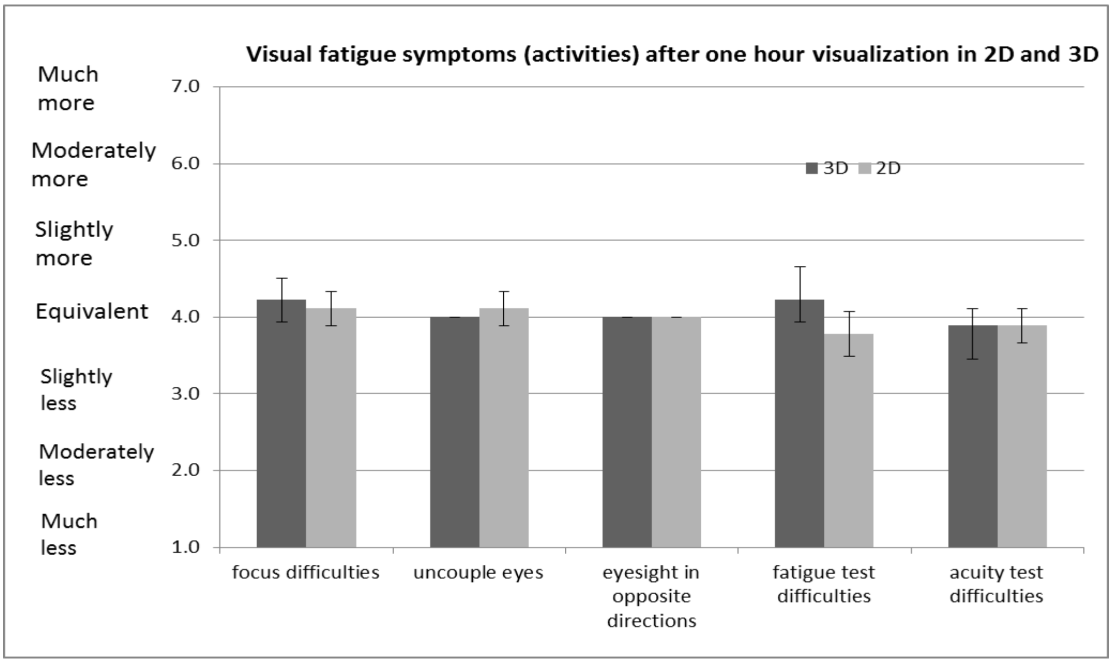


Figure 4-11 : Visual fatigue symptoms (activities) after one hour visualization in 2D and 3D

4.4.3 EEG measurement

EEGLAB toolbox was used to do the further analysis of EEG noise-free data. For each viewer, there were two one-hour processed EEG data sets in 2D and 3D conditions, respectively. Common EEG data analysis is to analyze the power spectrum of EEG data. Thus, Fast Fourier Transform (FFT) function provided by EEGLAB was used to transform each one-hour EEG data sets from the voltage-time signal to voltage-frequency signal for each channel. There is one basic question we are particularly interested in:

- Are the EEG signal power spectrums different between 2D and 3D conditions?

In order to investigate this question, we compared the power spectrum of 2D and 3D conditions channel per channel for each viewer. The power spectrums of Fp2 channel for all the viewers are plotted as an example in Figure 4-12. In Fp2 channels, the power density of beta band for Viewers 1, 2, 3, 5, 6, 7 in 3D condition is higher than the 2D condition. The amplitude of difference varies depending on viewer and frequencies. For Viewer 8 and Viewer 9, the differences between power spectrums of 2D and 3D conditions in the beta band are very small. Only for Viewer 4, the power density of 2D in the beta band is higher than the power density of 3D. Further statistical analysis was made by EEGLAB “parametric” statistical function. Paired student T test were used to calculate statistical significance between 2D and 3D conditions for all the viewers per frequency. $P < 0.05$ is used as threshold of rejecting the null hypothesis. The result of the mean power spectrum of Fp2 and its statistical significance mark is presented in Figure 4-13. It confirms our observation that in beta band, the 3D condition’s power density is significantly higher than the 2D condition.

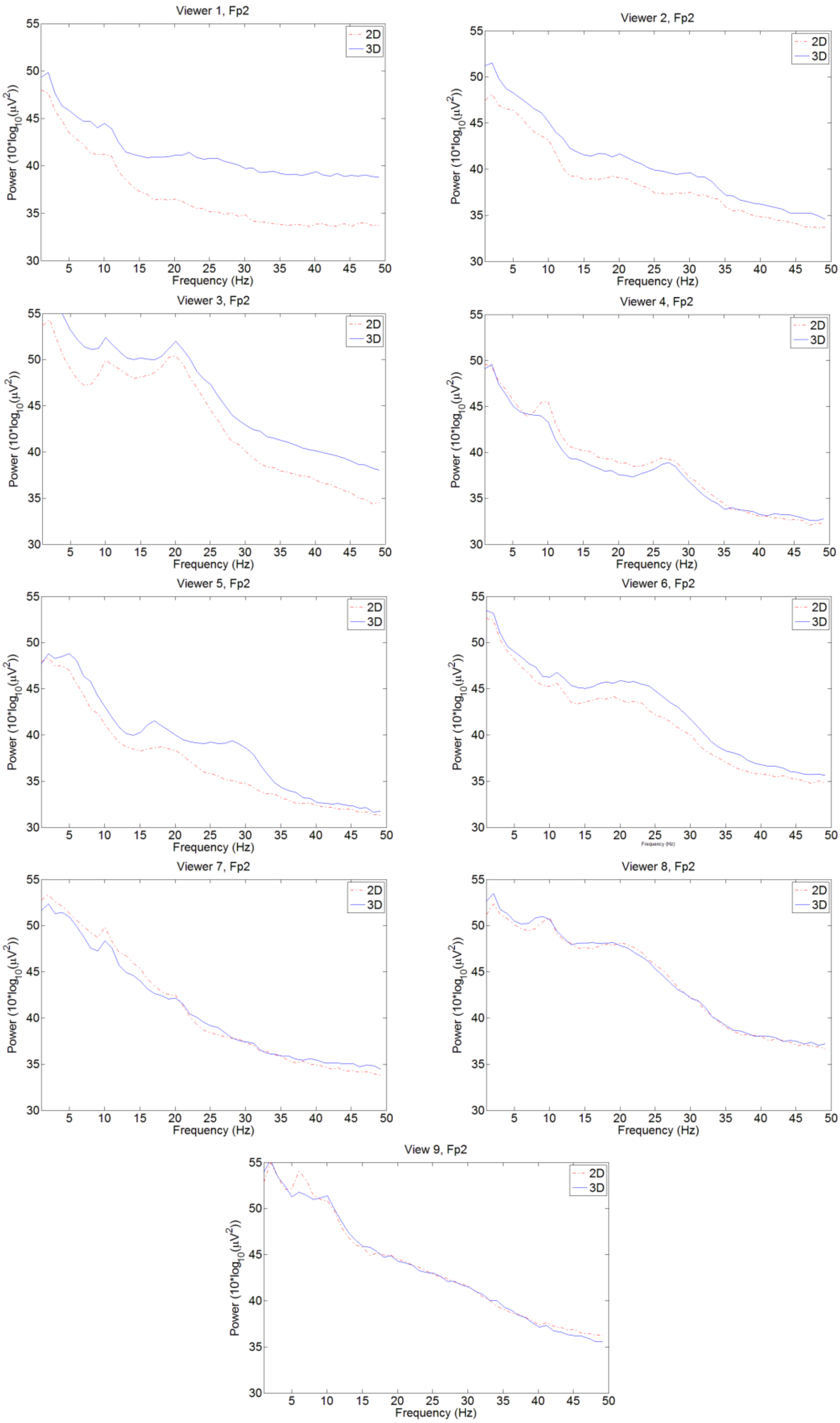


Figure 4-12 : Power spectrums of Fp2 channel for all the subjects

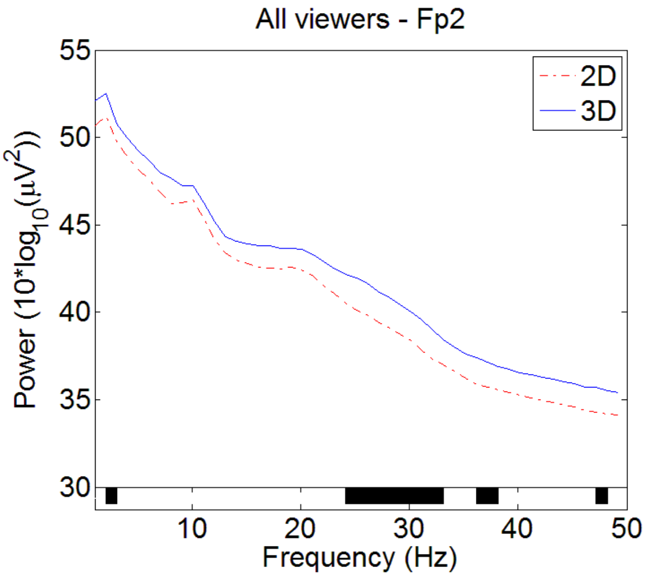


Figure 4-13 : Mean power spectrum of Fp2 channel for all viewers with statistical significance mark. (The bottom black bar indicates whether the mean value between 2D and 3D conditions at that frequency is significantly different on a 95% confidence level)

Figure 4-14, Figure 4-15 and Figure 4-16 plot the mean power spectrum with statistical significance marks for the EEG channels in Frontal Lobe, Temporal Lobe and central line, parietal lobe and occipital lobe, respectively. In frontal lobe, the significantly higher power densities for the 3D condition mainly occur in beta band of all five EEG channels and the gamma band of three EEG channels (F3, Fz, F4). In temporal lobe and central line, the curves of 3D and 2D power spectrum are quite similar and those are no statistically significant differences. In parietal lobe and occipital lobe, the significantly higher power densities occur mainly in gamma band for all the channels.

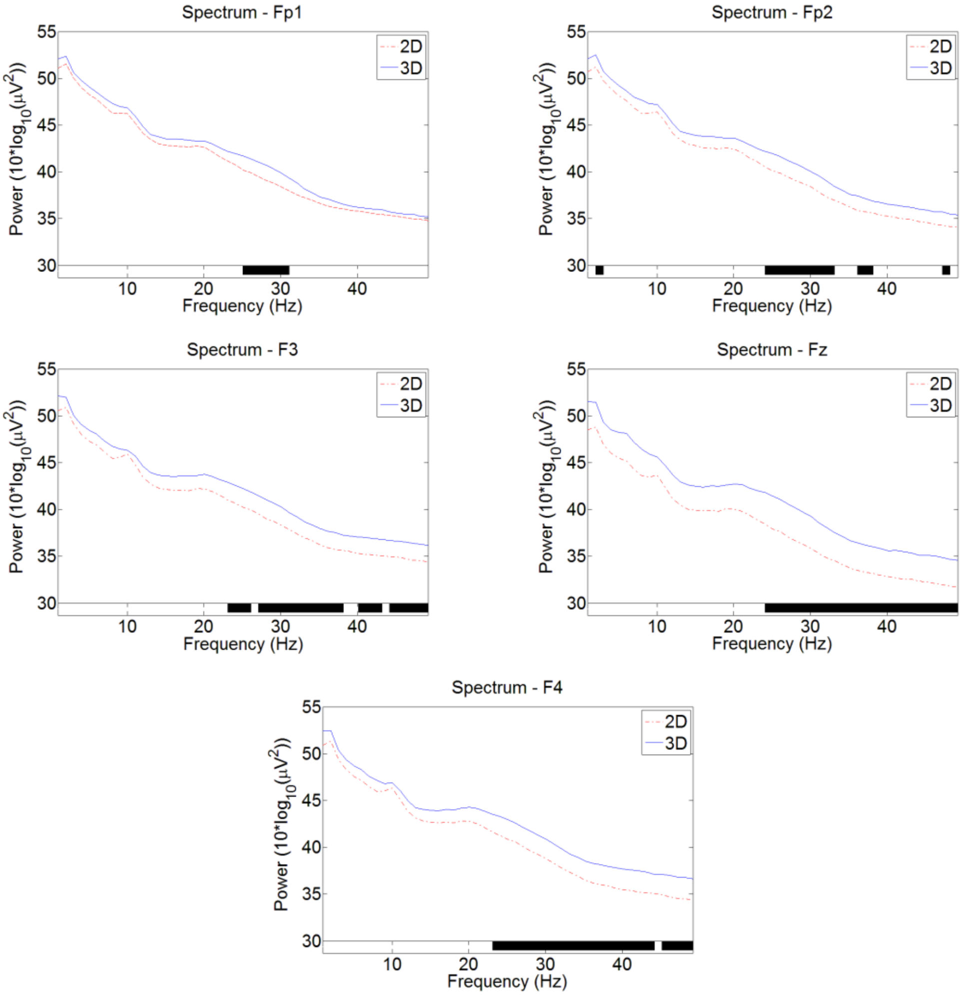


Figure 4-14 : Mean power spectrum with statistical significance mark for EEG channels in the frontal lobe

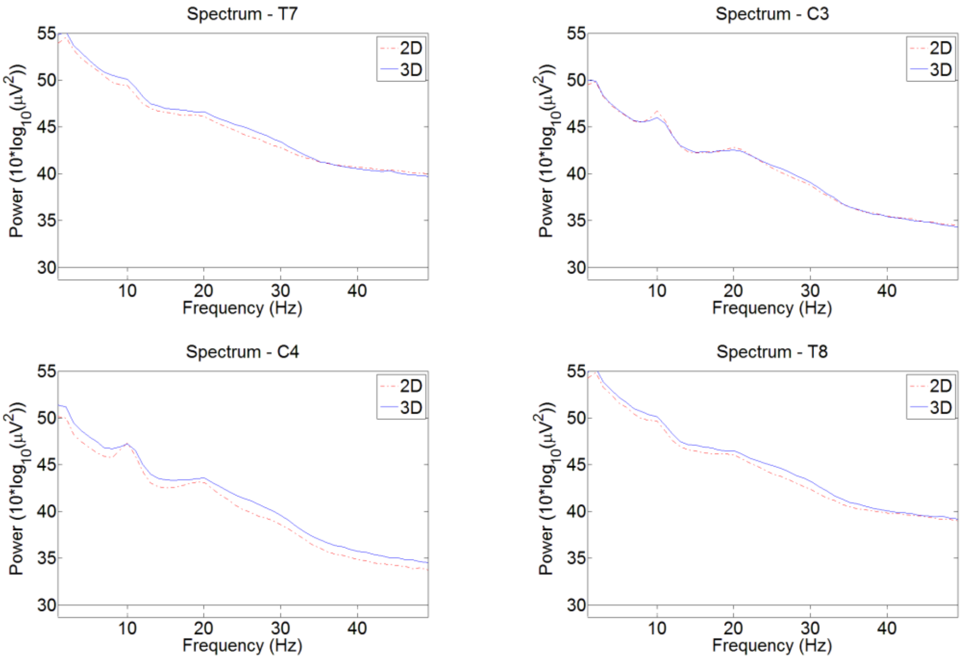


Figure 4-15 : Mean power spectrum with statistical significance mark for EEG channels in the temporal lobe and central line

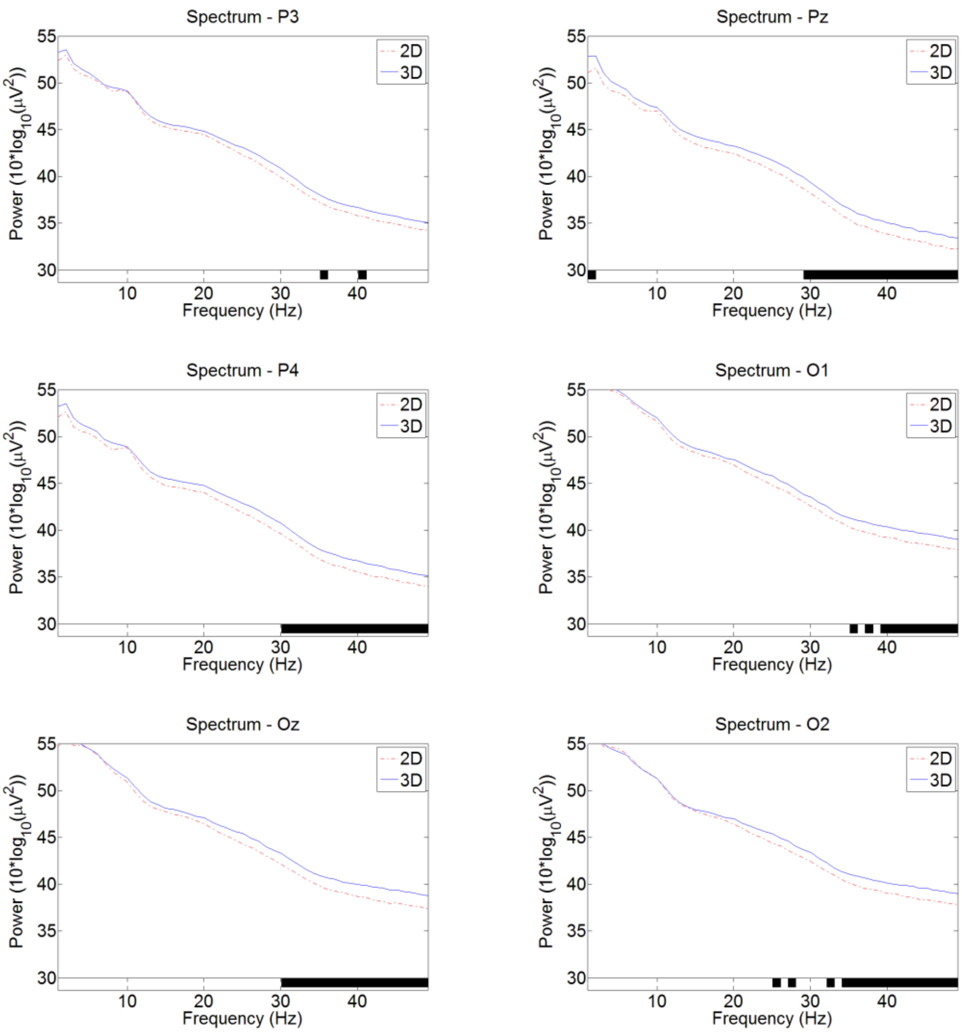


Figure 4-16: Mean power spectrum with statistical significance mark for EEG channels in the parietal lobe and occipital lobe

Furthermore, in order to analyze the temporal change of EEG signal, every one-hour EEG data set was segmented into 3 parts evenly: 0 to 20 minutes, 20 to 40 minutes and 40 to 60 minutes. For temporal statistical analysis, one-way repeated ANOVA (By EEGLAB statistical “parametric” function) method was used to calculate the significance of differences. There are no significant differences with respect to the temporal variation in both 2D and 3D conditions for most of the channels. However, by comparing the 2D and 3D conditions in difference temporal periods, it seems that the range of significant frequency bands tends to increase in the second period (20 to 40 minute) but to reduce in the third period (40 to 60 minute).

Figure 4-17 plots two examples as Fp1 (top) and Pz (bottom)’s mean power density in different time periods of 2D and 3D conditions with statistical significance analysis.

For Fp1 channel, only very narrow band around 10 HZ in the 2D condition is marked as significant change for the temporal variation. In the 3D condition, EEG power spectrum density is insignificant with respect to the increase of viewing duration. If comparing 2D and 3D conditions in different temporal periods, in 0-20 minutes the range of significant band locates only around 28HZ. It expands to be around 22 to 30HZ in 20 to 40 minutes. In the 40-60 minutes, no significant difference is found between 2D and 3D for Fp1 channels.

For the Pz channel, similar trends can be found in the Figure 4-17 (bottom). The range of significant bands is 16-50 HZ in 20 to 40 minutes. It is wider than only 36 to 50 HZ (partially) in 0 to 20 minutes and 30-50 HZ in 40-60 minutes.

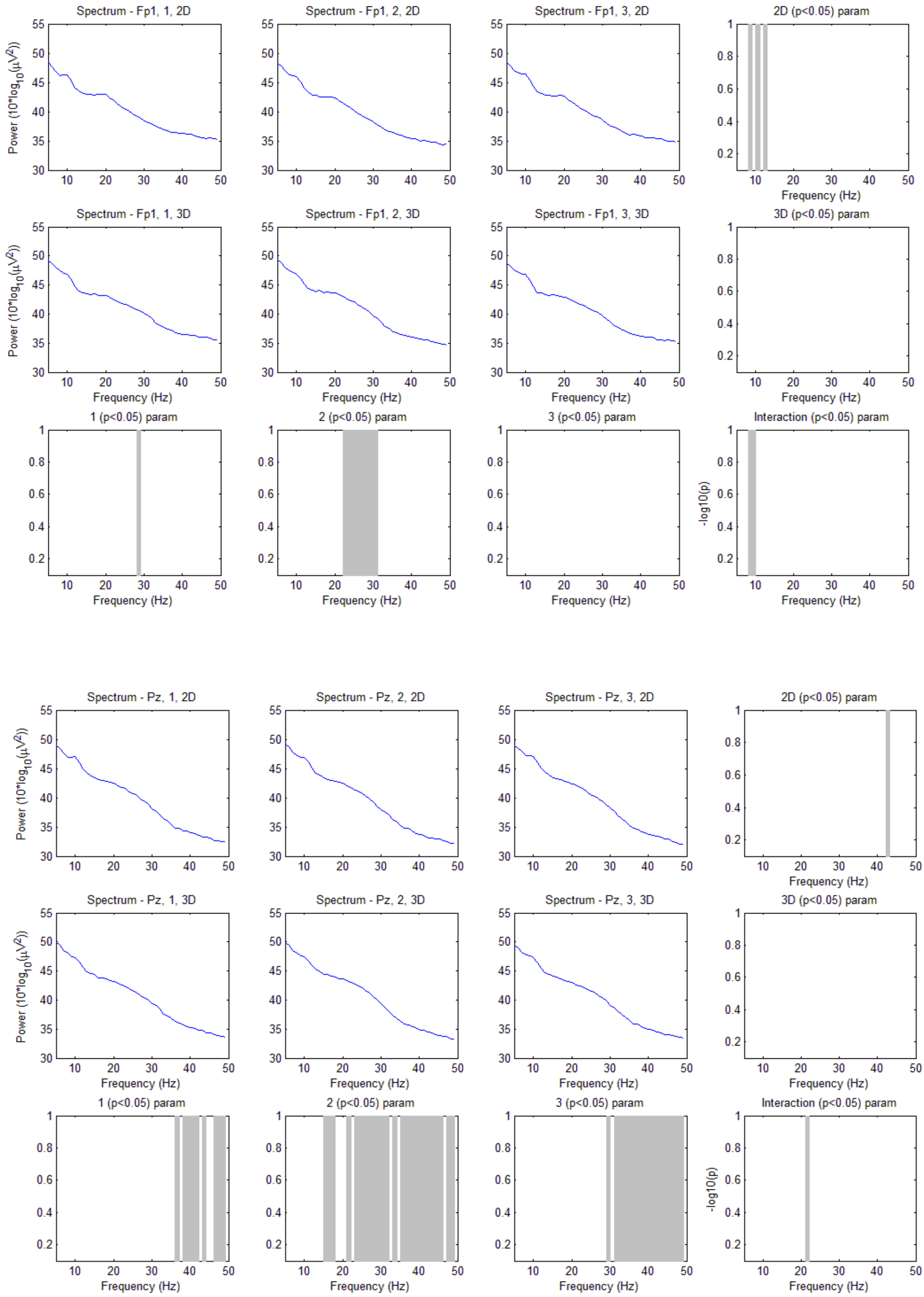


Figure 4-17 : Fp1 (top) and Pz (bottom)'s mean power spectrum in different time periods (1 as 0 to 20 minutes; 2 as 20 to 40 minutes; 3 as 40 to 60 minutes) of 2D and 3D conditions with statistical significance analysis

4.5 *Discussion*

The main findings of this chapter are summarized as follows:

- From the vision test and the questionnaire test, there were no significant evidences indicating that one hour of 3D viewing in this experiment caused more visual fatigue than 2D viewing.
- In most of the EEG channels located in the frontal, parietal and occipital lobes, the power of the beta or/and gamma bands were higher in the 3D condition rather than in the 2D condition.
- By segmenting the one hour EEG data into three parts (0-20 minutes, 20-40 minutes, 40-60 minutes), the EEG signals in both 2D and 3D conditions did not show significant change as viewing duration increased. However, by comparing the 2D and 3D power spectrum in the same time periods, the range of significant band where 3D has higher power density than 2D, tended to increase in the second period (20 to 40 minutes) but to reduce in the third period (40 to 60 minutes).

Compared to the previous research , Li et al. (Li et al., 2008) reported that:

- From subjective test, the visual fatigue level in 3D was higher than 2D.
- From EEG measurement, in most of the channels, the power of high frequency (>12 Hz) was stronger in the 3D condition rather than in the 2D condition and it tends to increase as presentation duration increased.

The main reason for the different results between this study and the previous study (Li et al., 2008), may be related to the content and viewing environment. In this study, the viewing environment and contents were selected to guarantee comfortable visualization. Our results may indicate that higher power strength in EEG spectrum in beta and gamma band is not necessarily related to visual fatigue. It can also be related to active concentration when people are more immersed into 3D content.

Thus, the conclusions drawn from this study are:

- If viewing environment and contents are optimized (i.e., content presented in comfortable viewing zone, no frequent pulse movement in depth, correct image asymmetries, minimize the crosstalk and etc.), viewing 3D does not necessarily result in visual fatigue.
- Concerning the brain activity of viewing 3D and 2D, there are some significant differences especially in the power strength of beta and gamma band in most of the frontal and posterior of cerebrum. However, it may not relate to visual fatigue.

Part II Impact of content acquisition on S-3DTV QoE

Chapter 5 New proposal of stereoscopic shooting rules to improve the QoE of S-3DTV

Contents

5.1 Introduction

As introduced in Chapter 1, for image acquisition and depth rendering, two main factors are assumed to affect the QoE of S-3DTV.

One is the stereoscopic distortion derived from geometry relationship between the camera spaces and the visualization space. It indicates the geometry difference between viewing stereoscopic images and actually viewing the real scene. To avoid stereoscopic distortion, the determination of shooting parameters should consider the scene parameters and the visualization parameters.

The other factor is the comfortable viewing zone which can be defined as limits of binocular fusion and depth of focus. It can avoid excessive binocular disparity and mismatch of convergence and accommodation. If the object in depth is outside of the comfortable viewing zone, viewing this object may induce visual discomfort, thus reducing the QoE. Generally, the comfortable viewing zone is defined in the final viewing environment. Thus, it should also be counted into constraints of visualization parameters.

To optimize the QoE of S-3DTV, these two factors should be taken into consideration. The objectives of the studies presented in this chapter are:

- To propose factors and thresholds to define precisely the stereoscopic distortion and the comfortable viewing zone;

- To propose rules to determine the shooting parameters based on the optimization of the stereoscopic distortion and the comfortable viewing zone;
- Design subjective QoE experiments to verify proposed shooting rules;

This chapter is organized as follows:

In Section 5.2, first, a geometrical model mapping the camera (parallel) space to the visualization space is presented; second, stereoscopic distortions including depth distortion factor and shape distortion factor are defined; third, comfortable viewing zone is defined; fourth, new stereoscopic shooting rules considering these two factors are proposed to determine shooting parameters in order to optimize the QoE of 3DTV.

In Section 5.3, an experiment is designed to judge the proposed stereoscopic shooting rules. Five synthetic scenes are produced. For each scene, five stimuli are generated in order to represent different levels of stereoscopic distortion and visual comfort. A subjective QoE assessment experiment with three QoE indicators (depth rendering, visual comfort and visual experience) is conducted to assess the stimuli. Results confirm that stimuli captured under the proposed shooting rules can ensure improved QoE.

5.2 New proposal of stereoscopic shooting rules based on stereoscopic distortion and comfortable viewing zone

A simple stereoscopic imaging system consists of image acquisition and visualization systems (see Figure 1-6). In image acquisition, binocular disparity information is recorded by two cameras with a horizontal shift as image disparity. In visualization, binocular depth information is represented by screen disparity which is a representation of image disparity in the S-3DTV display. When viewers watch stereoscopic images in front of the S-3DTV display, retinal disparity reflects the screen disparity. From the physical point of view, the above procedure is only a geometry mapping from one system to another system. Geometrical distortion is hence predictable. From the physiological point of view, viewing stereoscopic images in S-3DTV display is only an illusion. To guarantee the comfortable viewing experience, the final visualization of binocular depth needs to follow certain constraints as defined as a comfortable viewing zone in Section 1.4.2.

To enhance QoE of stereoscopic images, it requires the understanding of geometry of stereoscopic imaging system as well as the understanding of the physiological constraints of stereoscopic viewing in S-3DTV system. In this section, we aim to propose stereoscopic shooting rules to determine the shooting parameters to avoid stereoscopic distortion and guarantee comfortable viewing experience.

5.2.1 Geometry of the camera space and the visualization space

As light ray travels in straight lines, the functionalities of camera and screen are to record and represent light respectively. Thus, it is possible to use geometrical models to represent and predict the transmission of light ray on the camera space and the visualization space. In this section, we aim to present the geometry of camera space and visualization space.

As presented in Section 1.5.1, there are two possible configurations for stereoscopic two-camera system, i.e., the Toed-in and the parallel camera configuration. Compared to toed-in configuration, parallel camera configuration was proved to be able to maintain linearity during the conversion from real space to stereoscopic images. Thus, puppet theatre effect (Yamanoue et al., 2006) and vertical disparity (Woods et al., 1993) can be avoided. The main interest of toed-in camera configuration was that it does not require post-production shift or CCD shift to achieve image convergence. However, practically, in case of toed-in camera configuration, post production for correcting the geometrical distortion and vertical disparity are still required to achieve the same quality as parallel camera configuration.

In this study, only the parallel camera condition is considered. The simplified geometry of the parallel camera space and the visualization (planar display) space from (Woods et al., 1993) are shown in Figure 5-1 and Figure 5-2, respectively. The following variables are used to derive the geometry model mapping the camera space to the final visualization space.

Camera spaces (Figure 5-1):

Camera field of view (degree): depends on the focal length and the CCD (charge-coupled device) size of the camera.

Convergence point (meter): for parallel camera configuration, the virtual convergence point is achieve by a shift in CCD or a post-production shift to create a virtual convergence plane (or zero disparity plane)

f (meter) - Focal length of the camera.

b (meter) - Inter Camera baseline: the distance between the first nodal points of the two camera lenses.

w_{ccd}, h_{ccd} (meter) - Camera sensor width and height: the horizontal and vertical size of the camera CCD sensor.

d_{cov}(meter) - Convergence distance: the distance from the virtual convergence plane (or zero disparity plane) to the camera focal length plane.

$shift_{ccd}$ (meter) - Sensor shift or post-production shift: is the distance by which the center of each image senor has been moved away (outwards) from the optical axis of the lens to achieve the convergence.

$$\boldsymbol{shift_{ccd}} = \frac{\boldsymbol{fb}}{\boldsymbol{2d_{cov}}} \quad \textbf{(5-1)}$$

(x, y, z) (meter) - The location of a point in the camera space (in front of the camera).

$(x_{CL}, y_{CL}), (x_{CR}, y_{CR})$ (meter) - The point in the physical space maps into the camera sensor, the location of left and right image points in respective camera sensors.

Visualization space (Figure 5-2):

W_{screen} H_{screen} (meter)- Screen Width and Height: the horizontal and vertical size of the screen.

V - Viewing distance: The distance from the observer’s eye to display plane.

B (meter) - Inter-pupil baseline: The distance between the observers’ left eye and right eye, for adults, typically 65mm.

$(X_{SL}, Y_{SL}), (X_{SR}, Y_{SR})$ (meter) - The location of the left and right image points in the display plane.

(X, Y) (meter) - The location of the point in visualization space, as stereoscopically viewed by the observer when displayed on the screen.

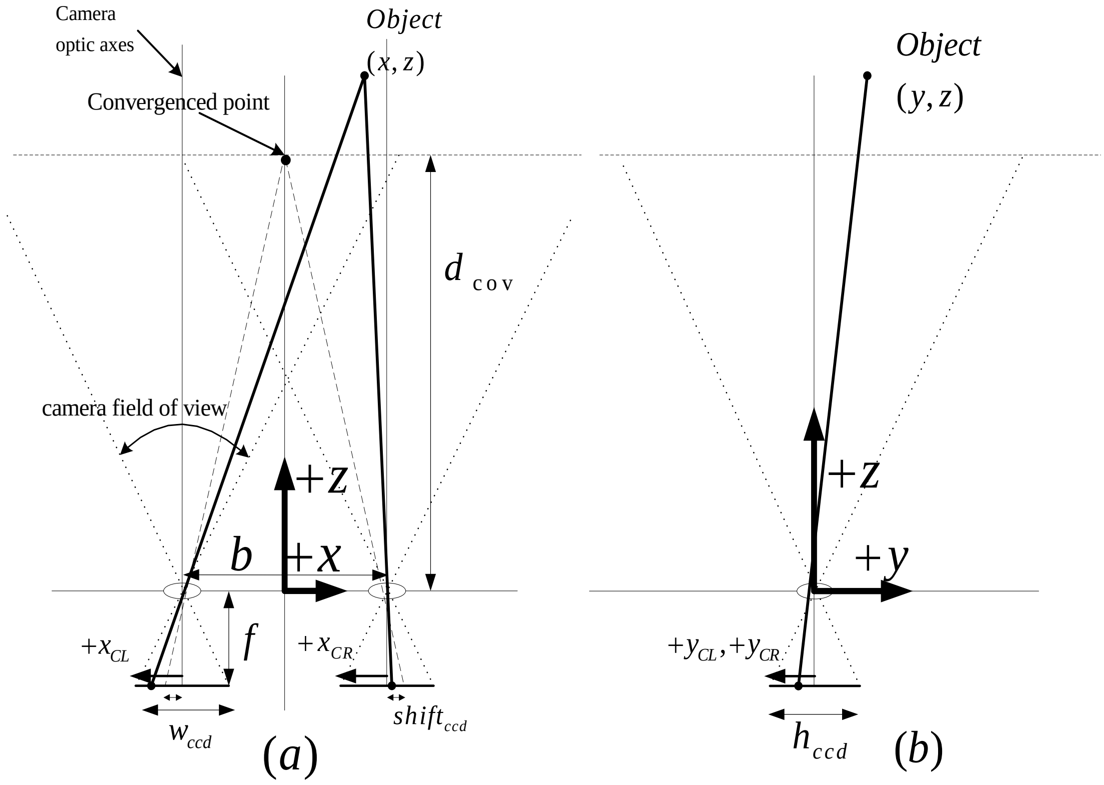


Figure 5-1 : Geometry of Parallel camera space (a) xz plane view (b) yz plane view

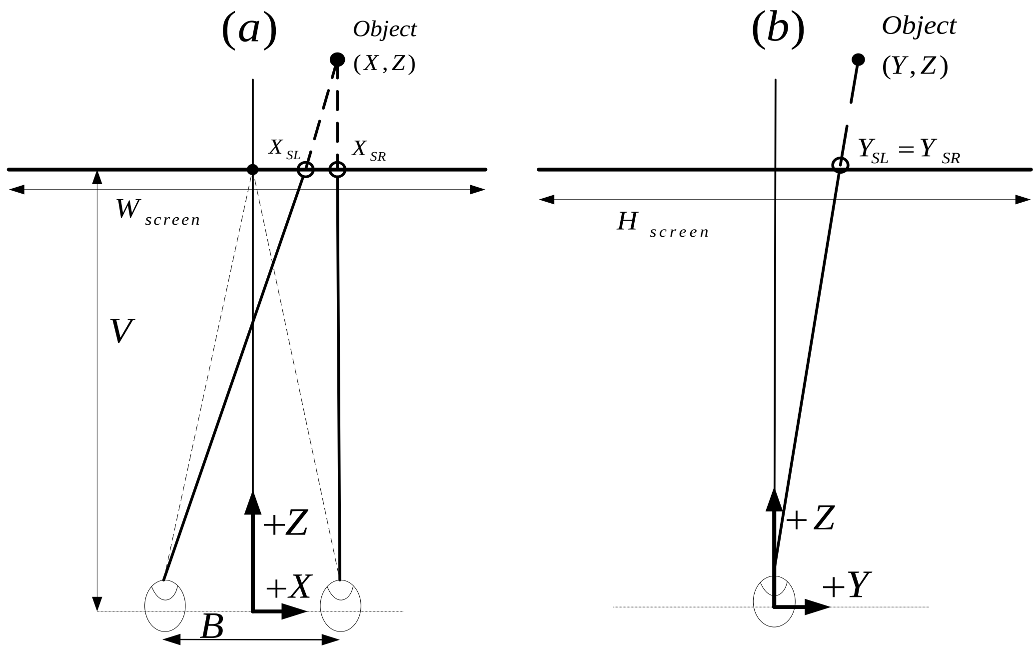


Figure 5-2 : Geometry of visualization space (a) XZ plane view (b) YZ plane view

Other system variables:

M - Frame Magnification factor, $M = W_{screen}/w_{ccd}$

The geometrical relationship of stereoscopic video systems can be mapped as follows:

$$Camera\ space\ (x, y, z) \rightarrow CCD\ sensor(x_{CL}, y_{CL})(x_{CR}, y_{CR}) \rightarrow Display\ plane(X_{SL}, Y_{SL})(X_{SR}, Y_{SR}) \rightarrow Visualzation\ space(X, Y, Z)$$

The overall transform from the camera space to the visualization space can be summarized as the below equations (derived from (Woods et al., 1993)) :

$$\boldsymbol{X = \frac{MBfx}{Bz+Mfb(1-\frac{z}{d_{cov}})}} \qquad \textbf{(5-2)}$$

$$\boldsymbol{Y = \frac{MBfy}{Bz+Mfb(1-\frac{z}{d_{cov}})}} \qquad \textbf{(5-3)}$$

$$\boldsymbol{Z = \frac{VBz}{Bz+Mfb(1-\frac{z}{d_{cov}})}} \qquad \textbf{(5-4)}$$

5.2.2 Stereoscopic distortion

Woods et al in (Woods et al., 1993) presented a number of stereoscopic distortions including depth plane curvature, depth non-linearity, shear distortion, depth and size magnification, key stone distortion and lens distortion. However, they did not define quantitative indicators for stereoscopic distortions. Yamanoue et al. in (Yamanoue et al., 2006) defined the reproduction magnification of the image as real size in the shooting space of the object divided by its apparent size in the stereoscopic image space. Their results showed that the difference of reproduction magnification between foreground and background can be used as the objective indicator for puppet theatre effect. Jones et al in (Jones et al., 2001) also presented their work to manipulate the shooting parameters in order to control the perceived depth in stereoscopic images. Holliman et al. in (Holliman, 2004b) mapped the depth in the real scene into three parts separately: near region, region of interest and far region. They proposed an algorithm to improve the perceived depth in region of interest to avoid shape distortion compared to other regions of the scene. However, their algorithm can only be applied for synthetic content creation.

In this study, we focus on stereoscopic depth distortion and shape distortion. The local depth variation around depth plane z can be represented by D_z as the derivative of the Z in the final visualization space with respect to z in camera space. It can be derived from equation (5-4) as follows:

$$\boldsymbol{D_z = \frac{dZ}{dz} = \frac{VBMfb}{\left[Bz+Mfb\left(1-\frac{z}{d_{cov}}\right)\right]^2}} \qquad \textbf{(5-5)}$$

D_z is used to represent the perceived stereoscopic depth distortion around depth plane z. It is a function of shooting parameters, visualization parameters and depth plane z. If D_z does not equal to one, it indicates that one unit of change of Z in visualization space does not correspond to one unit of change of z in camera space. In this case, viewers will not perceive the same binocular depth by viewing stereoscopic images compared with actually viewing the real scene.

Similarly, D_x and D_y are the derivatives of the X and Y in the final visualization space with respect to x and y. They can be derived from equation (5-2) and (5-3) as follows:

$$\boldsymbol{D_x} = \frac{\boldsymbol{dX}}{\boldsymbol{dx}} = \frac{\boldsymbol{MBf}}{\boldsymbol{Bz}+\boldsymbol{Mfb}(\boldsymbol{1}-\frac{\boldsymbol{z}}{\boldsymbol{d_{cov}}})} \qquad \textbf{(5-6)}$$

$$\boldsymbol{D_y} = \frac{\boldsymbol{dY}}{\boldsymbol{dy}} = \frac{\boldsymbol{Mfb}}{\boldsymbol{Bz}+\boldsymbol{Mfb}(\boldsymbol{1}-\frac{\boldsymbol{z}}{\boldsymbol{d_{cov}}})} \qquad \textbf{(5-7)}$$

They are used to represent the local variations of image magnification (2D size) in x and y axes, respectively. Because in most of cases D_x equals to D_y, for simplicity, only D_x will be used in the following to represent the image magnification.

Furthermore, by combining the local variation of depth and 2D size, people can perceive the shape of an object. It is also important to maintain the shape consistency between camera space and visualization space. Thus, a new factor representing the 3D shape distortion is defined as D_s. It denotes the change ratio of D_z versus D_y as follows:

$$\boldsymbol{D_s} = \frac{\boldsymbol{D_z}}{\boldsymbol{D_x}} = \frac{\boldsymbol{Vb}}{\boldsymbol{Bz}+\boldsymbol{Mfb}(\boldsymbol{1}-\frac{\boldsymbol{z}}{\boldsymbol{d_{cov}}})} \qquad \textbf{(5-8)}$$

Similar to the stereoscopic depth distortion indicator D_z, D_s is also a function of the shooting parameters, the visualization parameters and the depth plane z. When D_s equals to one, 3D shape around depth plane z in the visualization space is maintained equal to the camera space. When D_s does not equal to one, stereoscopic shape distortion occurs. For example, a cube may be perceived as a cuboid and a Round object may be perceived as oval object in case of distortion in stereoscopic shape. In the following of the thesis, stereoscopic 3D shape distortion factor D_s is used as the main indicator of the stereoscopic distortion.

Assuming that the parameters of the visualization space are known and constant, changing the parameters of image acquisition will change the stereoscopic depth and shape distortion. However, practically, there are different camera models in parallel configuration allowing different degrees of freedom to camera parameters. In the following, analysis of the stereoscopic distortions under different camera models is presented. The Full resolution 22 inch desktop display with 1680x1050 pixels resolution as presented in Table 3-3 is used as the default visualization display. The default viewing distance is three times of screen height.

5.2.2.1 Orthostereoscopic model

Diner in (Diner, 1991) introduced Orthostereoscopic as “A 3D image is orthostereoscopic when it perfectly replicates human vision”. The conversion ratio of the camera space to the visualization space in case of orthostereoscopic system is constantly one to one. Thus, there are no stereoscopic distortions.

This model is achieved by capturing with a focal length that perfectly matches the human angular field when replicated in the visualization environment. This means:

$$X = x, Y = y, Z = z$$

Combining the above condition with equations (5-2, 5-3, and 5-4), we can derive the condition below:

$$b = B, f = \frac{d_{cov}}{M}, V = d_{cov}$$

Thus, in the orthostereoscopic setting condition, three conditions need to be fulfilled:

(1) Camera baseline equals to inter pupil baseline,

(2) Focal length equals to the convergence distance divided by the frame magnification factor.

(3) Viewing condition should be fixed to as the convergence distance.

Figure 5-3 depicts stereoscopic distortions analysis in orthostereoscopic condition. There is no stereoscopic distortion existing in this condition, i.e., the stereoscopic images under the orthostereoscopic condition represent perfectly the physical spaces in geometry to human eyes. However, due to strict requirement for focal length and viewing condition, this special model only applies in science application, medical application and military robot application which have very strict requirement for representation of the real world.

$f = 0.076\ b = 0.065\ d_{cov} = 1\ w_{ccd} = 0.035\ W_{screen} = 0.046\ V = 1\ B = 0.065\ (meter)$

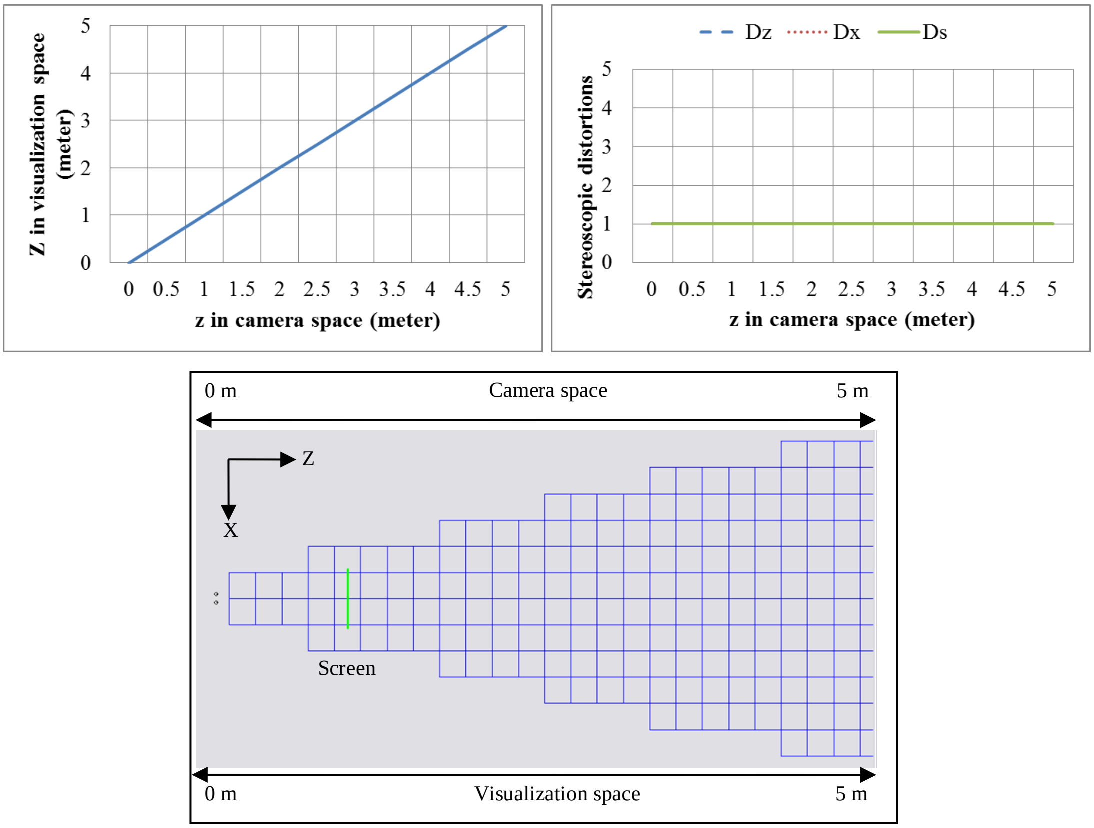


Figure 5-3 : Stereoscopic distortion in the case of orthostereoscopic system (top left) plot of Z in visualization space versus z in camera space (top right) plot of stereoscopic distortion versus z in camera space (bottom) illustration of the shape distortion in visualization space (each rectangle in camera space has been arbitrarily chosen to be 0.2 meters long in x axis and 0.2 meters long in z axis)

5.2.2.2 Fixed camera baseline model

Fixed camera baseline model is a more practical model compared with the Orthostereoscopic model. It is mostly used in the consumer stereoscopic camera system. For example, Fujifilm W1 has a fixed camera baseline in 65mm. Its focus length ranges from 6.3mm to 18.9mm (1/2.3 inch CCD) which is equivalent to 35-

105mm focal length on a 35mm camera. The advantage of the fixed camera baseline model is that the camera positions can be pre-calibrated to avoid view asymmetries.

The common setting in this model is that camera baseline equals to inter pupil baseline so that:

$$b = B$$

By inserting the above condition into equations (5-5, 5-6 and 5-7), we can get

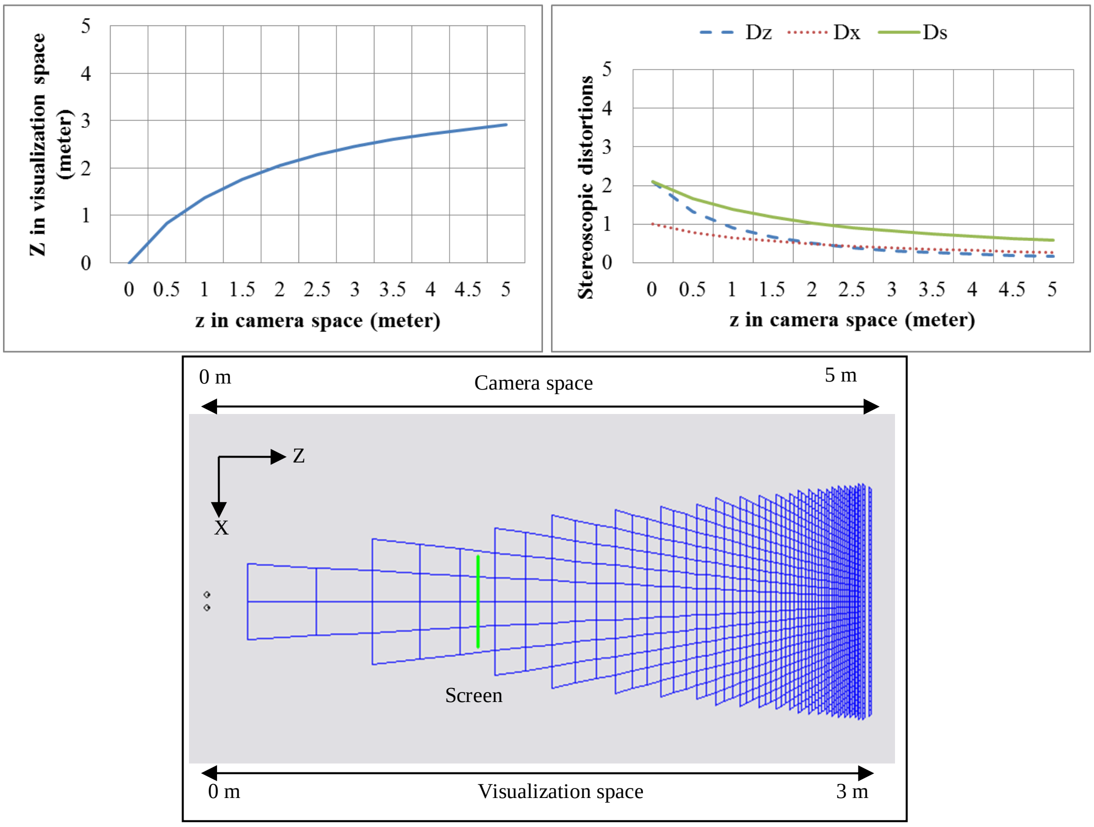
$$\boldsymbol{D_x = D_y = \frac{Mf}{Bz+Mf(1-\frac{z}{d_{cov}})}} \qquad \textbf{(5-9)}$$

$$\boldsymbol{D_z = \frac{VMfb}{\left\{B\left[z+Mf\left(1-\frac{z}{d_{cov}}\right)\right]\right\}^2}} \qquad \textbf{(5-10)}$$

In most of the stereoscopic cameras using a fixed camera baseline, the change of focal length is possible and it also results in changing the final depth rendering.

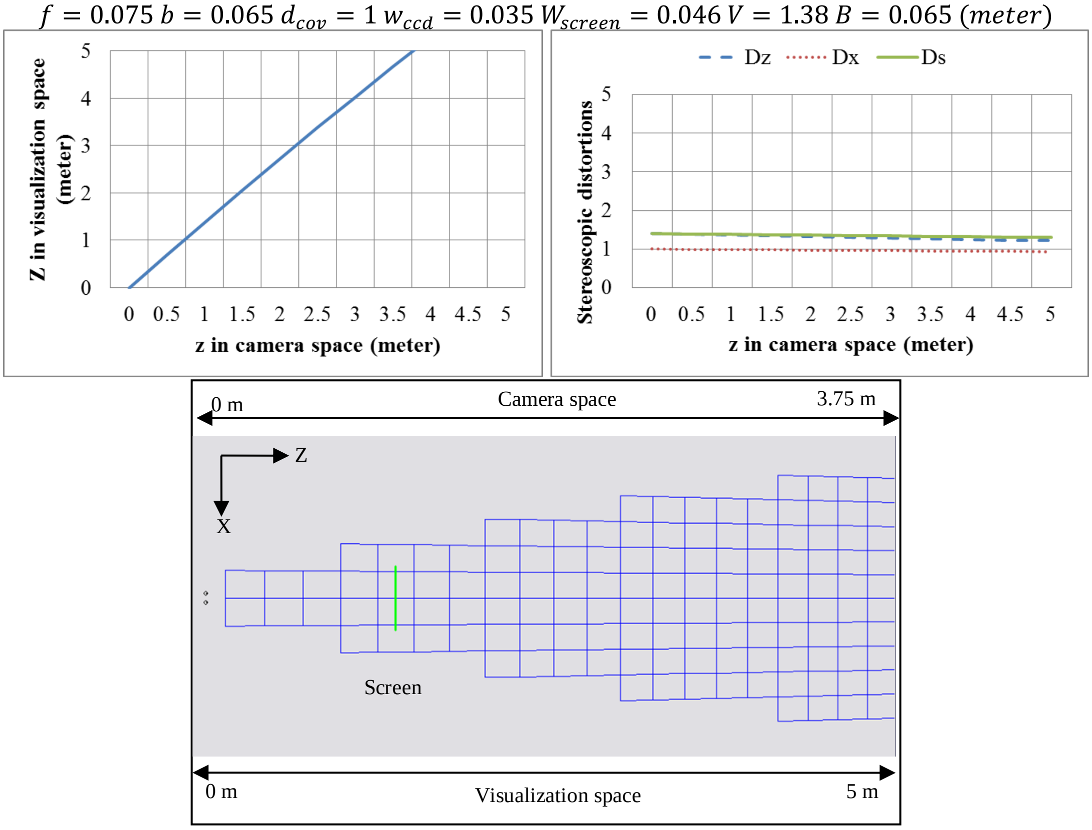
Figure 5-4 plots the depth rending analysis of a “full frame” (35mm) camera sensor, fixed focal length camera model with a 50mm (a) and 75mm (b) focal length values. The stereoscopic distortion in the 50mm focal length condition is not linear and the values reduce by the increase of distance in depth. In the near distance (< 2 meters in camera space), stereoscopic shape distortion factors are larger than 1 so that the shape of objects are stretched in depth direction. In far distance (> 2 meters), stereoscopic distortion factors are smaller than 1 so that the shape of objects are compressed in the depth direction. For the 75 mm case, the distortion factors are nearly linear and constant. However, the depth space is stretched, e.g., 1 meter in physical depth distance is perceived in 1.4 meter in visualization space.

$$f = 0.05\ b = 0.065\ d_{cov} = 1\ w_{ccd} = 0.035\ W_{screen} = 0.046\ V = 1.38\ B = 0.065\ (meter)$$



(a)

$$f = 0.075\ b = 0.065\ d_{cov} = 1\ w_{ccd} = 0.035\ W_{screen} = 0.046\ V = 1.38\ B = 0.065\ (meter)$$



(b)

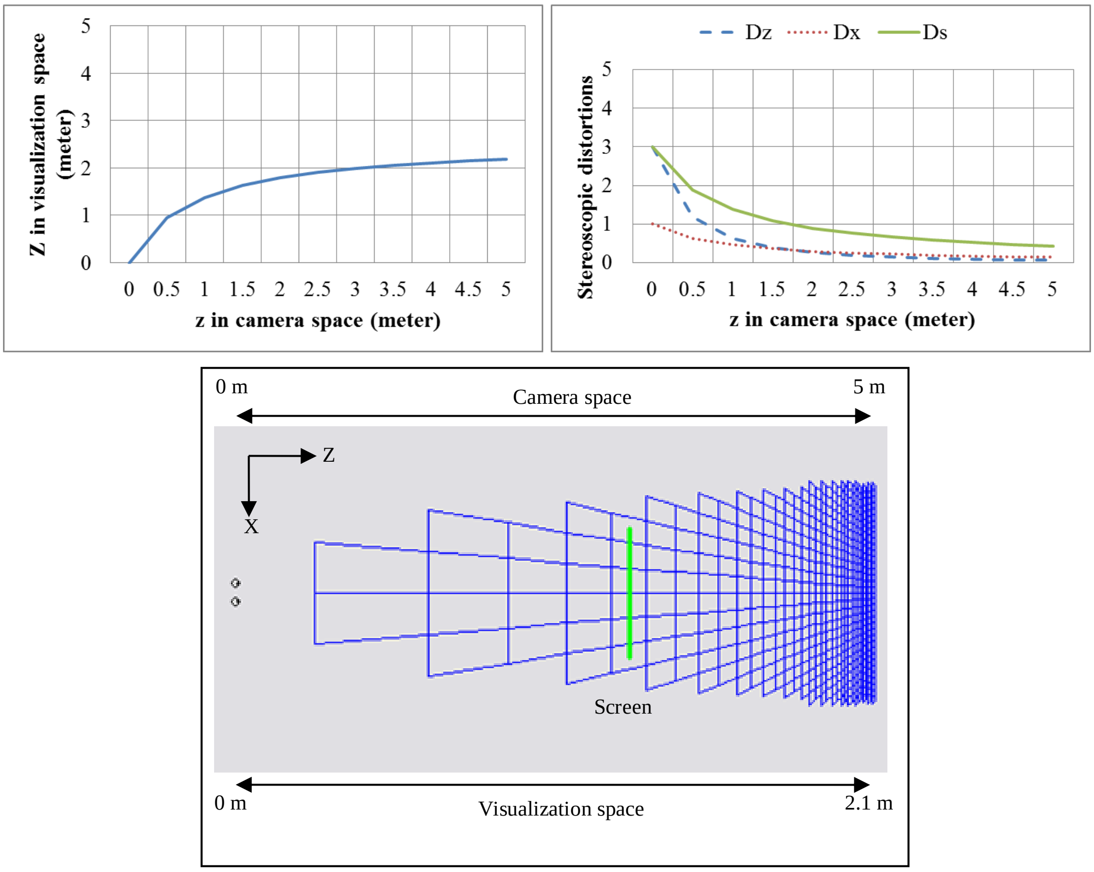
Figure 5-4 : Stereoscopic distortion in the case of fixed camera baseline system
(a) 50mm focal length; (b) 75mm focal length

5.2.2.3 Fixed focal length model

Fixed focal length lenses normally can provide better optic performance than adaptable focal length lenses. Thus, they are widely used in professional shooting especially to guarantee the image quality, e.g., movie shooting. Furthermore, due to the complexities and difficulties to synchronize the focal length between the left and right camera in stereoscopic two-camera system, fixed focal length models are also widely used in stereoscopic movie production. In this case, it is still possible to change the camera baseline to affect the depth rendering.

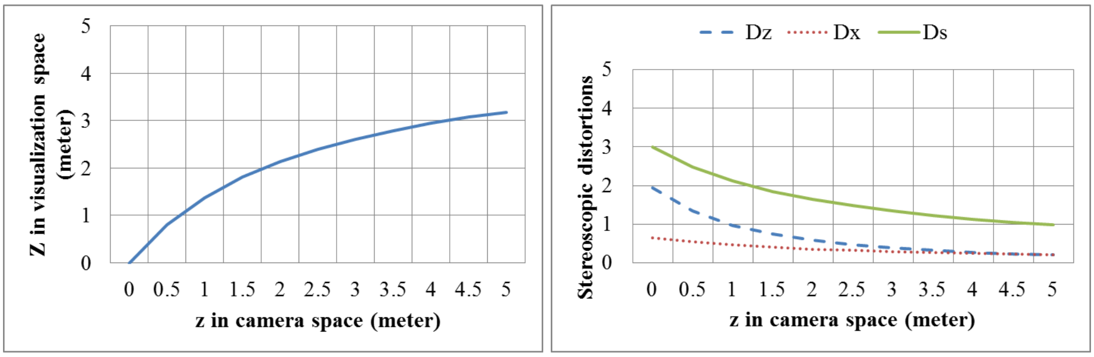
Figure 5-5 depicts how the adaptation of the camera baseline (65mm, 100mm and 140mm) affects the final depth rendering. The larger camera baseline is used; the more stretched depth space is achieved in visualization space.

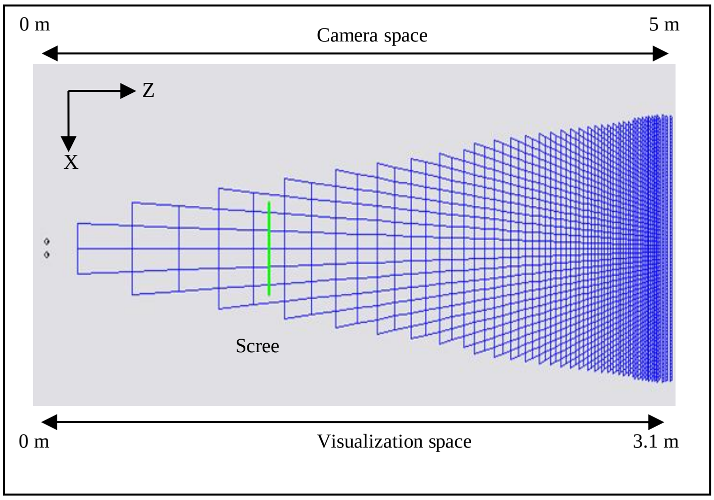
$$f = 0.035 \; b = 0.065 \; d_{cov} = 1 \;\; w_{ccd} = 0.035 \; W_{screen} = 0.046 \; V = 1.38 \; B = 0.065 \; (meter)$$



(a)

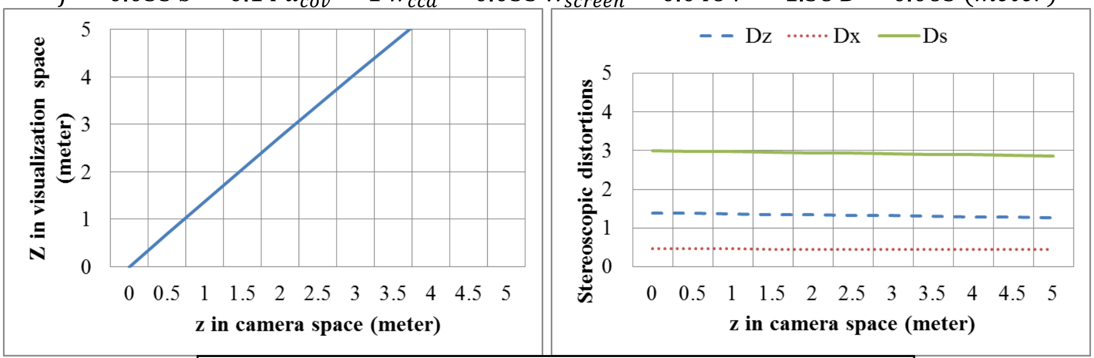
$$f = 0.035 \; b = 1 \; d_{cov} = 1 \; w_{ccd} = 0.035 \; W_{screen} = 0.046 \; V = 1.38 \; B = 0.065 \; (meter)$$

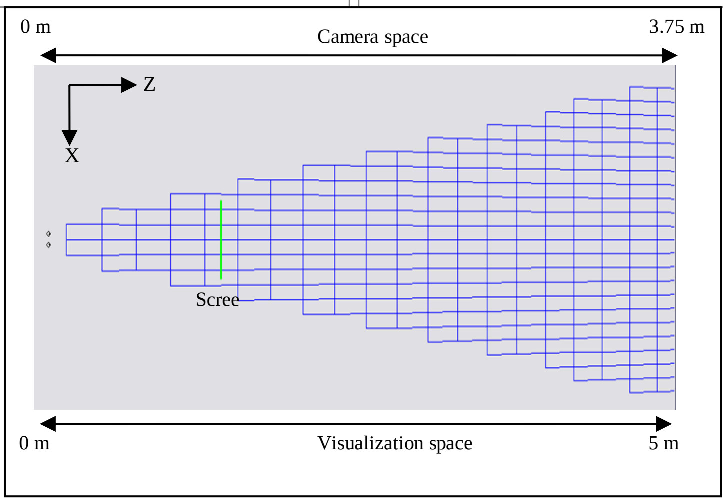




(b)

$f = 0.035 \; b = 0.14 \; d_{cov} = 1 \; w_{ccd} = 0.035 \; W_{screen} = 0.046 \; V = 1.38 \; B = 0.065 \; (meter)$





(c)

Figure 5-5 : Stereoscopic distortion in the case of fixed focal length system (a) 65mm camera baseline; (b) 100mm camera baseline; (c) 140mm camera baseline

5.2.3 Comfortable viewing zone

As defined in (Lambooij et al., 2007), the comfortable viewing zone is a perceptual range where binocular fusion is possible and blur is not perceived so that stereoscopic visual comfort should be maintained. In Section 3.3, the comfortable viewing zone is assumed to be ±0.2 diopters, to facilitate the analysis of depth rendering ability of 3D displays. In this section, a more thorough discussion of the comfortable viewing zone is presented.

For visual comfort, earlier in the last 90s, S. Pastoor(Pastoor, 1992) discussed about the human factors e.g. disparity range and concluded that visual comfort for stereoscopic video systems is a key factor related to its success of competing with 2D systems. In (Wöpking, 1992), the author proposed the visual comfort threshold of 70 arcmin for disparities based on subjective assessment. 0.3 diopters (reciprocal value of distance) and 60arcmin were suggested in (Lambooij et al., 2007) as limits of Depth of focus (DOF) and binocular disparity, respectively. ITU-R BT.1438 (ITU, 2000) recommends the threshold from Hiruma and Fukuda‘s work (Broadbent, 2004) that the stereoscopic pictures are suggested to be displayed within the depth of field of the human eye which is ±0.3 diopters to avoid defocusing of image. Yano et el. in (Yano et al., 2004) suggested a more conservative value as ±0.2 diopters since visual discomfort was clearly induced when images were displayed outside the corresponding range of depth of focus, and even within this range, visual discomfort can be induced if the image were moved in depth according to a step pulse function. In (Kooi and Toet, 2004), the authors suggested 2 to 3 PD (prismatic diopter) for threshold of horizontal disparities.

In professional stereoscopic shooting activities, the $1/30^{th}$ rule of thumb of 3D (Mendiburu, 2009) is suggested and widely used in stereo photography to avoid excessive disparities. It stipulates that the inter-axial distance should be not more than $1/30^{th}$ of the distance from the camera to the first foreground object. However, it is an empirical method and only can contain a rough estimation and suggestion for camera parameters. It does not cover the feature of possible variation of screen size and viewing distance. Thus, for cinema shooting, this rule is suggested to be adapted to $1/100^{th}$ and for very short lenses $1/10^{th}$ may be used.

The above proposed threshold and rules related to the comfortable viewing zone are summarized in Table 5-1.

Table 5-1 : Summary of the studies related to the comfortable viewing zone

Studies	Method	Threshold
(Wöpking, 1992)	Subjective experiment for visual comfort	70 arcmin for binocular disparity
(Lambooij et al., 2007)	Summary of literature theory	±0.3 diopters for DOF and 60 arcmin for binocular disparity
(Broadbent, 2004) (ITU, 2000)	Measure the accommodation response to stereoscopic TV images	±0.3 diopters for DOF
(Yano et al., 2004)	Subjective experiment for visual comfort	±0.2 diopters for binocular disparity
(Mendiburu, 2009)	Empirical suggestion	$1/30^{th}$ rule to decide the camera baseline
(Kooi and Toet, 2004)	Subjective experiment for visual comfort	2 to 3PD (60 to 90 arcmin) for binocular disparity

Most proposed limits of the comfortable viewing zone are functions of the viewing distance. The range of the comfortable viewing zone increases as the viewing distance increase. Figure 5-6 plots the limits of the comfortable viewing zone following the proposed thresholds shown in Table 5-1.

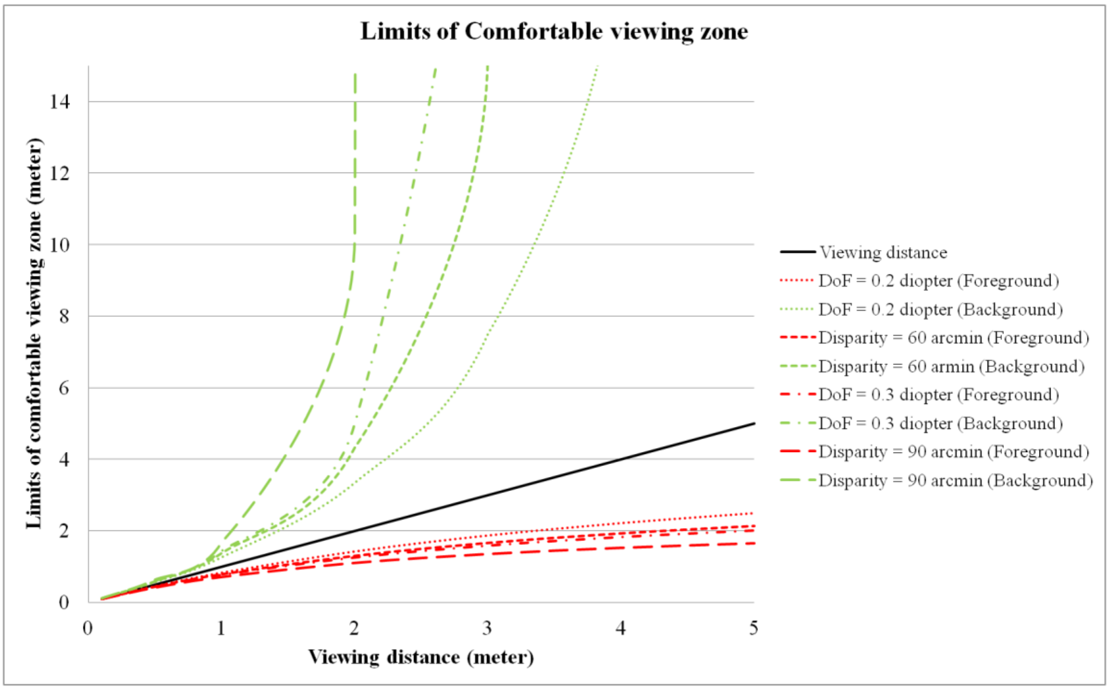


Figure 5-6 : Limits of the comfortable viewing zone

Combining the proposals in the literature to ensure visual comfort for watching S-3DTV, the most conservative value ±0.2 diopters as illustrated in Figure 5-7 is assumed to be a general limit for the comfortable viewing zone in this study. Further experiments in Section 5.3 and Chapter 6 will confirm this threshold.

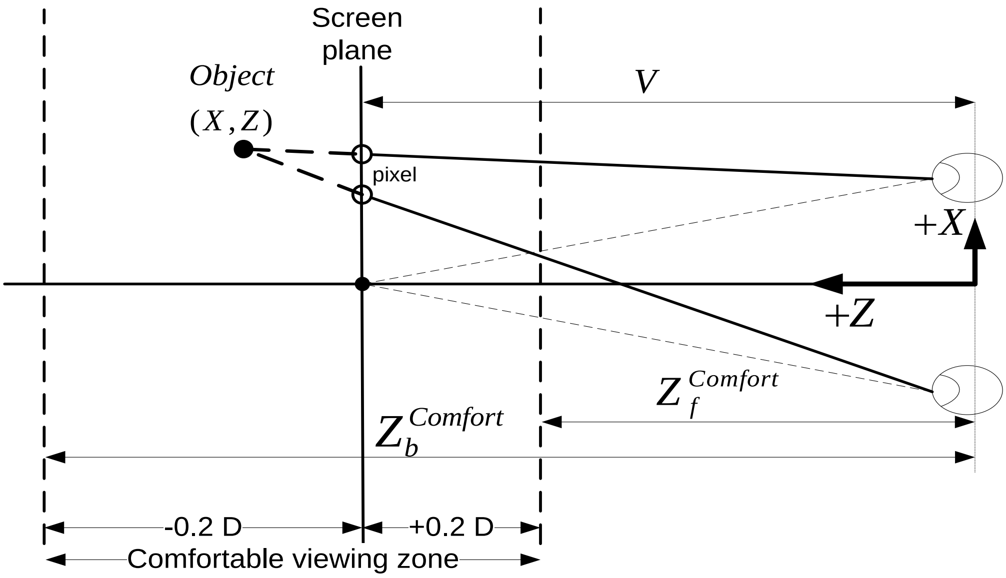


Figure 5-7 : The comfortable viewing zone ($DoF = 0.2$ $diopter$)

5.2.4 Improved stereoscopic shooting rules

In order to optimizing the stereoscopic shooting, by considering the effect of stereoscopic distortion and comfortable viewing zone, three shooting rules are proposed in this section.

Shooting Rule 1

From equation (5-8), the stereoscopic shape distortion factor D_s is not a linear function, only in some extreme cases, e.g. orthostereoscopic as discussed in the last session. D_s is a constant of one when the parameters of the camera space and the visualization space can fulfill:

$$if\ b = B, f = \frac{d_{cov}}{M}, V = d_{cov}, then\ D_s = 1$$

However, in practical application, the above conditions are hard to fulfill. For instance, stereographers are used to select their own camera focal length based on the field of view of the camera. Viewing distance is normally depending on the final display size and resolution. Thus, there are a lot of constraints considering the camera parameter and the visualization parameters which result in the difficulties to keep the perceived depth space to be linear.

The Region of Interest (ROI) is a selected subset of samples within a dataset identified for a particular purpose. This concept is often used in the image and video processing in order to do conditional optimization. For example, ROI functionality is provided in the JPEG 2000 standard to give a desirable encoding for the ROI in the image (Andrew, 2010). For stereoscopic content production, Holliman in (Holliman, 2004b) also proposed a method to improve the depth perception in the ROI. Regarding the difficulties of maintaining the linear shape distortion for stereoscopic images, optimization of stereoscopic image acquisition should target the ROI in depth as priority. Hence, z_{RoI} is defined as the depth plane at which the ROI object locates. The basic improved shooting Rule 1 can be defined as follows:

(1) Adapt the changeable parameters to guarantee the shape distortion factors D_s^{RoI} (D_s in the plane of ROI as z_{RoI}) to approximate one as much as possible.

Shooting Rule 2

The comfortable viewing zone based on the limit of depth of focus (diopters) is defined as follows: $Z_f^{Comfort}$ is the limit of the absolute foreground distance in the visualization space and $Z_b^{Comfort}$ is the limit of the absolute background distance. They can be derived as follows:

$$\boldsymbol{Z_f^{Comfort} = 1/\left(\frac{1}{V} + DoF\right)} \qquad \textbf{(5-11)}$$

$$\boldsymbol{Z_b^{Comfort} = \begin{cases} 1/\left(\frac{1}{V} + DoF\right), if\ V < DoF^{-1} \\ \infty, if\ V \geq DoF^{-1} \end{cases}} \qquad \textbf{(5-12)}$$

where V represents viewing distance (see Figure 5-7). Knowing scene depth range as $(z_f^{scene}, z_b^{scene})$ in camera space, the perceived depth range in final visualization can be computed as $(Z_f^{scene}, Z_b^{scene})$ by Equation (5-4). Visual comfort can be guaranteed only if

$$\boldsymbol{Z_f^{Comfort} < Z_f^{scene} < Z_b^{Comfort}\ and\ Z_f^{Comfort} < Z_b^{scene} < Z_b^{Comfort}} \qquad \textbf{(5-13)}$$

There are two methods to fulfill the equation (5-13)'s condition. The first method assumes that the camera parameters and visualization parameters are known and unchangeable, only the scene parameters are adaptable. From equation (5-4), we can derive its inverse function:

$$\boldsymbol{z^{camera}} = \frac{\boldsymbol{Mfbd_{cov}}}{\boldsymbol{Bd_{cov}\left(\frac{V}{z^{visualization}}-1\right)+Mfb}} \qquad \textbf{(5-14)}$$

Based on this equation, we can get $(Z_f^{Comfort}, Z_b^{Comfort}) \rightarrow (z_f^{Comfort}, z_b^{Comfort})$. Thus, the visual comfort condition can be fulfilled only if the scene range can be limited within the comfortable depth zone in the camera space as

$$\boldsymbol{z_f^{Comfort} < z_f^{scene} < z_b^{Comfort} \;\; and \;\; z_f^{Comfort} < z_b^{scene} < z_b^{Comfort}} \qquad \textbf{(5-15)}$$

However, in most cases, the director or stereographers are imposing the freedom to design the scene range freely. So the second method is to adapt the camera parameters (e.g., focal length, convergence distance and camera baseline) to guarantee the perceived scene range to locate within the comfortable viewing zone as Equation (5-13). The basic improved shooting rule 2 can be summarized as follows:

(2) Guarantee that the perceived scene range $(Z_f^{scene}, Z_b^{scene})$ is **maintained within the comfortable viewing zone** $(Z_f^{Comfort}, Z_b^{Comfort})$ **by adapting the scene parameters or camera parameters.**

Shooting Rule 3

However, in some scenarios, the above two basic improved rules cannot assemble each other so that priority should be decided. We assume that the visual comfort problem is more important than the stereoscopic shape distortion in usability oriented applications (television broadcasting, movie and etc.). In this thesis, the shooting rule is mainly designed for usability oriented application so that the combination and priority of the two basic improved rules is defined as follows:

(3) When Rule 1 and Rule 2 cannot be fulfilled simultaneously, Rule 2 is prior to Rule 1 which means the visual comfort is more important than the stereoscopic shape distortion.

However, in utility oriented application, e.g., medical and space science stereoscopic viewing, the strategy of priority may be different.

5.3 Verification of the proposed improved shooting rules

In the previous section, stereoscopic shooting rules were proposed to avoid stereoscopic distortion and guarantee comfortable viewing. However, all the proposals are based on theoretical analyses or assumptions. Their perceptual impact on the QoE of S-3DTV remains to be confirmed. In this section, we design a subjective QoE experiment to investigate the perceptual impact of the proposed stereoscopic shooting rules.

5.3.1 Stereoscopic image (synthetic) generation

For verification of the proposed improved shooting rules, practical stereoscopic content acquisition or generation is required. In order to avoid the view asymmetry problems such as camera misalignment and colorimetric, synthetic stereoscopic content generation was chosen. Five scenes representing different depth ranges were designed by manipulating the materials from an open source animation project "Big buck bunny" (BlenderFoundation, 2008). The final images were rendered by Blender software (BlenderFoundation, 2010).

5.3.1.1 Stereoscopic scene categorization and selection

Table 5-2 : The stereoscopic scene categorization

Name	Micro space	Personal space	Action space	Vista space
Depth range	<1m	1-3m	3-15m	>15m

In the paper (E.Cutting and M.Vishton, 1995), the depth discrimination function delimits three types of space around the observer - personal space, action space and vista space – each served by different sources of depth cues and with different sensitivities. In this experiment, we divided the original personal space (0-3 meter) into micro space (<1 meter) and personal space (1-3 meter) in order to precisely distinguish the stereoscopic scene. Then stereoscopic scenes can be categorized by depth range as described in Table 5-2. Based on the above categorization, five scenes from the "Big buck bunny" content as shown in Figure 5-8 were chosen in order to cover all the categories. The purpose of selecting scenes in different depth range is to judge:

- Whether the depth range (i.e., located in different space) of the scene has potential impact on the QoE?
- Whether the proposed shooting rules can improve the depth perception for scenes in different depth range?

All the rendering settings for each scene were carefully adapted in order to provide the sharpest image in different depth layers, i.e., the original blur effect in the background was removed in order to avoid unnatural blur (Lambooij et al., 2009a). The description of these five scenes is shown in the Table 5-3. ROI objects were selected and defined carefully by experts in order to coincide with the viewer's interest in the image. As default, the converged plane or so-called zero disparity plane was set to be in the center of the object to make the final rendered object located on or around the display plane in order to get it more comfortable and sharp.

Table 5-3 : Overview of the five scenes and their characteristics

Scene Number	z_b^{scene} Near plane (m)	z_f^{scene} Far plane(m)	z_{RoI}^{scene} ROI plane(m)	Depth space	Description
1	0.33	2.4	0.56	Micro space	Bunny holding the arrow and bow
2	1.9	17	4.7	Action space	Bunny walking from the tree to flower
3	5	10	5	Action space	Robe skipping bunny
4	1.1	10	1.5	Personal space	Three standing squirrels
5	20	52	30	Vista space	A tree and the background forest

Figure 5-8 : Five selected scenes from "Big buck bunny" (top left: scene 1; top right: scene 2; mid left: scene 3; mid right: scene 4; bottom: scene 5 as defined in Table 5-5)

5.3.1.2 Acquisition and Post processing

All the images were rendered by Blender software in the resolution of 1920x1080 pixels. Multisampling (8 samples) anti-aliasing was used to make the edges smooth and all the addition blur effect was disabled to guarantee natural sharpness. The virtual camera inside Blender is 32mm x 16mm size sensor. A special python plug-in of Blender was developed in order to implement the stereoscopic shooting in the Blender software. Two individual parallel virtual cameras with controlled camera parameters were used to capture the scene. In order to avoid the black border after the post production shift, extended borders related to the converged shift were rendered for every image.

The post production included the post-production shift to generate the convergence plane located at the screen plane and stereoscopic format conversion (see Appendix B. Representation format conversion) to generate the compatible format for the S-3DTV screen.

5.3.1.3 Selected Camera parameters

StereoCalculator Software was developed in order to ease the selection of camera parameters by implementing the improved shooting rules. The final visualization display was the Hyundai S465D line-interleaved display (1920x1080 pixels) with polarized glasses. The viewing distance was computed to fulfill the 1 minute of arc visual acuity threshold of pixel. Camera sensor was the default Blender camera sensor (32mm x 16mm) and the camera focal length was fixed in each scene. The converged distance was set to equal to the center of the ROI plane. The only adaptable camera parameter for depth rendering was the camera baseline. The fixed camera parameters are shown in Table 5-4.

Table 5-4 : Fixed camera parameters

f(mm)	$w_{ccd} \times h_{ccd}$ (mm)	d_{cov}	$W_{screen} \times H_{screen}$ (m)	V (m)	B (mm)
24.65(Scene 1), 28.59(Scene 2), 35(Scene 3-5)	32 x 16	z_{RoI}	1.01 x 0.57	1.82	65

For each scene, five different conditions were defined in order to justify the proposed improved shooting rule:

- Condition (1): 2D image, the left view from the stereoscopic image pair was used directly as the 2D image;
- Condition (2): *DOF* equals to 0.1 D;
- Condition (3): *DOF* equals to 0.2 D, the most conservative limit from literature proposals for maintaining visual comfort (see Figure 5-7);
- Condition (4): *DOF* equals to 0.3 D, which is suggested by the ITU-R BT.1438 as the threshold of depth of focus to maintain visual comfort;
- Condition (5): D_s^{RoI} equals to 1, i.e. there is no shape distortion in the ROI plane.

For condition (2) to (4), each DOF value represents different levels of the comfortable viewing zone ($Z_f^{Comfort}$,$Z_b^{Comfort}$) (higher value means larger depth range) as well as different levels of shape distortion. Moreover, the final perceived depth fulfills the requirement from both the equation (5-13) and the below equation:

$$(\boldsymbol{Z_f^{scene}} = \boldsymbol{Z_f^{Comfort}}) \ \textbf{or/ and} \ (\boldsymbol{Z_b^{scene}} = \boldsymbol{Z_b^{Comfort}}) \qquad \textbf{(5-16)}$$

which guarantees that the perceived scene range is located in the comfortable viewing zone defined by different DOF values by enlarging or compressing the perceived depth. The condition (5) is following the basic Rule 1 but without considering Rule 2.

Camera baselines were calculated to fulfill the above five conditions for five selected scenes as shown in the Table 5-5 as well as the calculated D_s^{RoI} :

Table 5-5 : Camera baseline and shape distortion of five different conditions in five scenes

Scene Number	2D	$DoF = 0.1$		$DoF = 0.2$		$DoF = 0.3$		$D_s^{RoI} = 1$	
		b (mm)	D_s^{RoI}	b (mm)	D_s^{RoI}	b (mm)	D_s^{RoI}	b (mm)	D_s^{RoI}
1	0	11	0.55	22	1.1	33	1.65	***20***	***1***
2	0	43	0.26	***85***	***0.5***	145	0.86	168	1
3	0	106	0.59	213	1.20	319	1.79	***178***	***1***
4	0	17	0.31	***37***	***0.69***	50	0.93	54	1
5	0	638	0.59	1283	1.2	1930	1.8	***1069***	***1***

The previous studies (Goldmann et al., 2010a, Goldmann et al., 2010c), only used the same group of 10, 20, 30,40, 50cm camera baselines for different scenes without considering the depth range differences of each scene. In this case, the same camera baseline in different scenes may not represent the same perceived depth level. In our study, DOF values can be used as the indicator of the normalized perceived depth range which can ease the comparison of different camera setting. The conditions which fulfill the proposed shooting rule in each scene are shown in bold and underlined in Table 5-5.

5.3.2 Subjective QoE assessment

In order to verify the proposed improved shooting rules, a subjective QoE assessment experiment was designed to investigate the effect of different scenes (five scenes as shown in Table 5-3) and test conditions (five different conditions as shown in Table 5-5) on three QoE indicators consisting of visual experience, depth rendering and visual comfort. The method of this experiment is presented as follows:

1) Stimuli: The test session was composed of five scenes as shown in Table 5-3. For each scene, five different stimuli had been considered corresponding to different conditions as shown in Table 5-5 so that there were overall 25 stimuli which were all still stereoscopic images.

2) Equipment: The subjective QoE assessment was conducted in a test room, which was compliant with the recommendations for subjective evaluation of visual data issued by ITU-R BT.500 (ITU, 2002). As mentioned previously, a 46 inch line-interleaved stereoscopic display with a native resolution of 1920x1080 pixels was used as the final visualization terminal. A digital video system (DVS) which can output 1920x1080 HD signal was used.

3) Observers: 28 observers were recruited to participate in this test. All of them were non-experts in the audiovisual and video domain. A vision test was performed on all testers to determine their visual performance and the potential impact on

results. The test includes monocular visual acuity test (distant vision acuity test, near vision acuity test), hyperopia trend, astigmatic trend, binocular distant vision acuity, fusion and stereoacuity. The vision test showed that all testers had a normal or corrected to normal visual acuity and a stereoacuity of < 1 minute of arc.

4) Procedure: written instructions detailing the task what they had to perform and the attribute they were asked to rate were given to the subjects before the start of the test. These instructions were then reiterated by the experimenter to ensure the observer understood the task. The SAMVIQ method was used in this test. Considering the 3D evaluation concepts and the test purpose, three QoE indicators (depth rendering, visual comfort and visual experience) as defined in Section 2.3.1 were used in this test. The whole test was separated into two experiments. Experiment I included a test session of depth rendering and a test session of visual comfort. Experiment II only included a test session of visual experience and it was organized in a different day in order to avoid the influence of the experiment I on experiment II.

   For each test session, five scenes, which have five stimuli in each scene, were evaluated by the subjects. For each scene, the subject could see all the five stimuli and report their perceptual opinion. These stimuli were shown as button A, B, C, D, and E. They can be viewed several times if the subject wished. The buttons were randomly reassigned to stimuli so that the subjects could not identify them. Each stimulus was shown with the duration of 7s and the subjects provided their score. Subjects were able to freely modify their score before the end of the test.

5) Outlier detection: The screening of subjects was performed according to the guidelines described in ITU-R BT.500 (ITU, 2002). For the Experiment I, 5 of the 28 subjects have been discarded as outlier. For Experiment II, 3 of 28 subjects have been discarded. Thus the final result analysis is based on the scores of the non-outlier subjects.

5.3.3 Result analysis

Experiment I

Figure 5-9 and Figure 5-10 plot the mean opinion scores and their confidence interval of depth rendering and visual comfort respectively in five conditions for the five scenes. The computed confidence interval is corresponding to a significance level of 95%.

Considering the depth rendering assessment, the results show that subjects can easily distinguish between the stereoscopic image and the 2D images. 2D is always scored "poor". It is not easy for subjects to distinguish the depth rendering between 3D in different perceived depth ranges. However, the condition (5) ($D_s^{ROI} = 1$) for shape optimization within the region of interest still shows slight advantage than the others in most of the scenes except the scene 2. By analyzing the scene setting and taking into account subjects' opinion about scenes composition, scene 2 has been identified as a special case. The windows violation was produced by the inappropriate depth position of the flower. The flower shown in scene 2 (see Figure 5-8) was perceived as floating in the foreground of the display in binocular depth perception. It contradicted to the pictorial depth cues from the image that the flowers was connected to the ground as well as the solid displayer border references in reality. It produced contradiction in higher level cognition activity of combining different depth cue, resulting in the bias of the results.

Concerning the visual comfort assessment, the results show that visual comfort decreases with the increase of binocular disparity. 2D condition is always ranked between “good” and “excellent”, better than all stereoscopic conditions. In most of the scenes, condition (2) ($DoF = 0.1$) and condition (3) ($DoF = 0.2$) are worse than the 2D condition, However, the visual comfort level is higher than 60 (good to excellent) except scene 2. The condition (4) ($DoF = 0.3$) presents a steeper reduction of visual comfort compared with the condition (3). It may suggest that 0.2 diopters is the threshold of comfortable viewing. Moreover, the QoE degradation with the increment of DOF value in scene 2 is steeper than the other scenes. It might indicate that depth cue contradiction produced more visual discomfort.

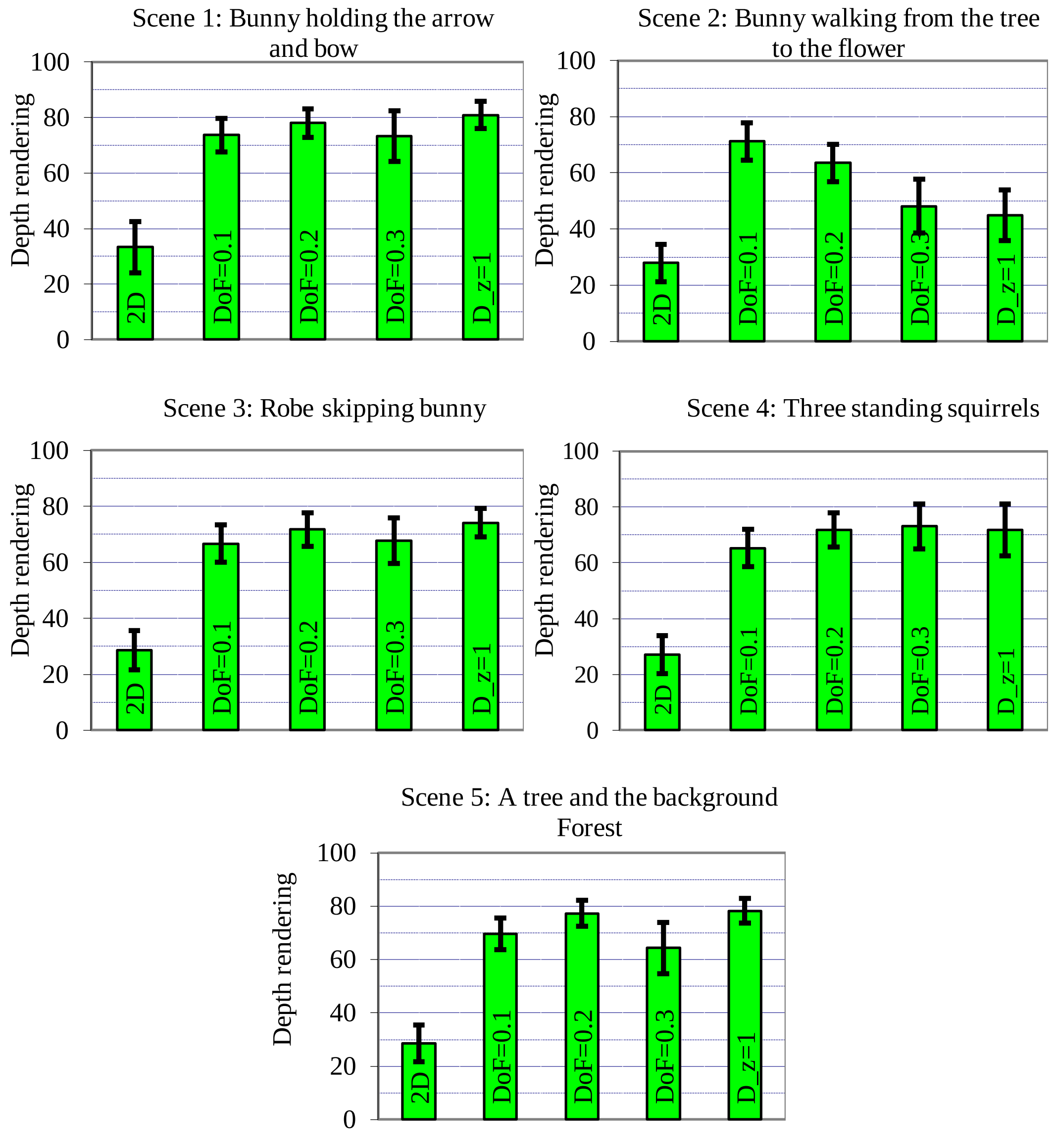


Figure 5-9 : Mean opinion scores and confidence intervals of depth rendering in five different conditions for the different scenes

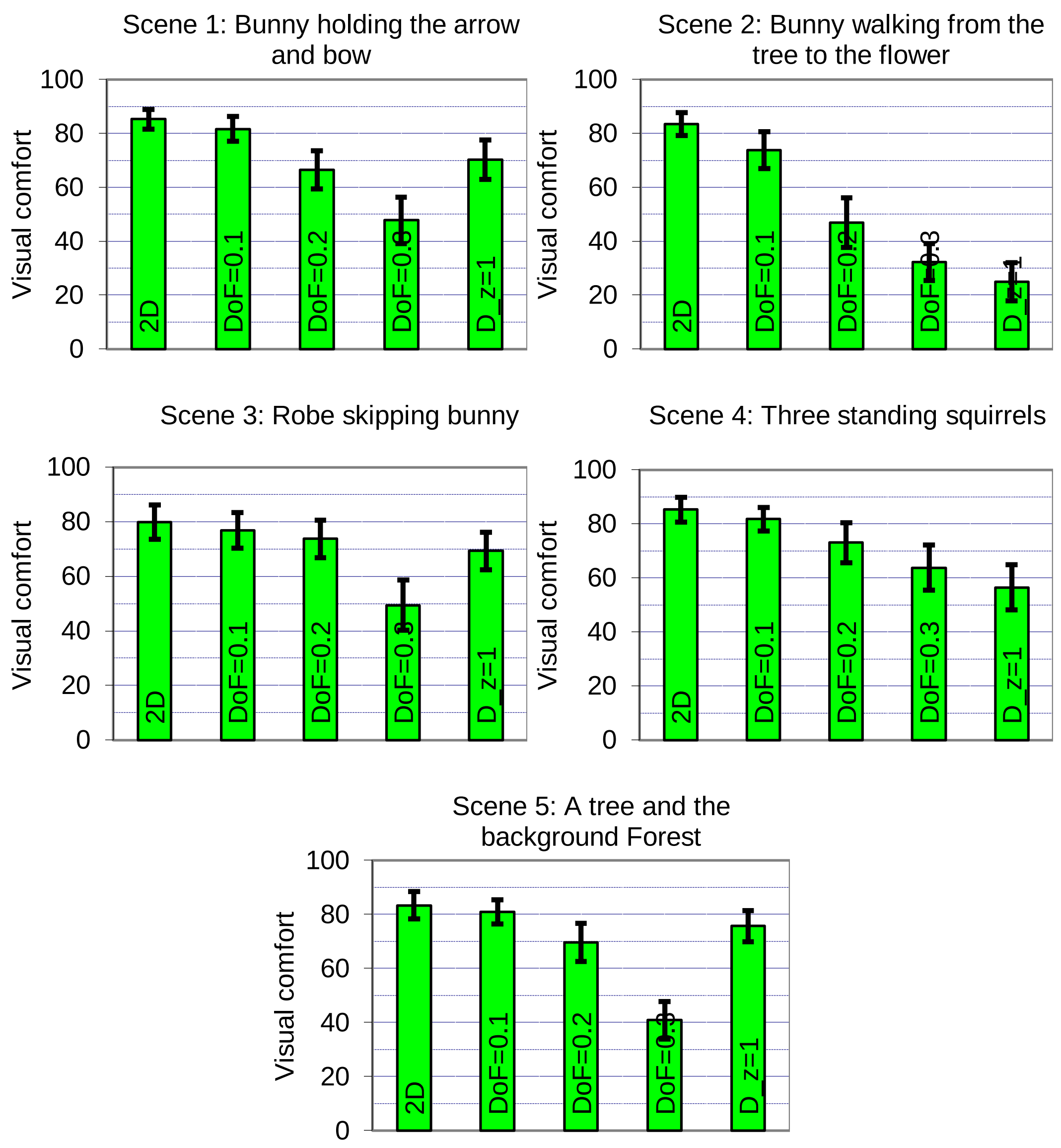


Figure 5-10 : Mean opinion scores and confidence intervals of visual comfort in five different conditions for the different scenes

Experiment II

Figure 5-11 plots the result of the visual experience test. Stereoscopic stimuli whose perceived depth ranges are within the comfortable viewing zone (0.2 diopters) are all scored above “good”. 2D images are scored around the level of “fair”, lower than the comfortable 3D images. The scores of condition (2) ($DoF = 0.1$) and condition (3) ($DoF = 0.2$) only have slight differences which are similar to the results of depth rendering. This might indicate that when visual comfort maintains “good”, the depth rendering is the dominant factors for visual experience. Scene 2 has additional visual discomfort problem due to depth cues contradiction. In this case, the ranking of visual experience scores in different condition is similar to visual comfort scores. This might indicate that visual comfort is the dominant factor of visual experience when visual discomfort matters. The selected conditions which fulfill the proposed shooting rules in each scene are marked in red/deep color in Figure 5-11. In Scene 1, Scene 3 and Scene 4, the selected conditions are rated as the highest scores. In Scene 2 and Scene 5, the selected conditions are rated as the second highest scores while the conditions

(2) are rated as the highest scores. This might indicate that in those scene, a more conservative depth of focus value ($DoF = 0.1$) should be suggested.

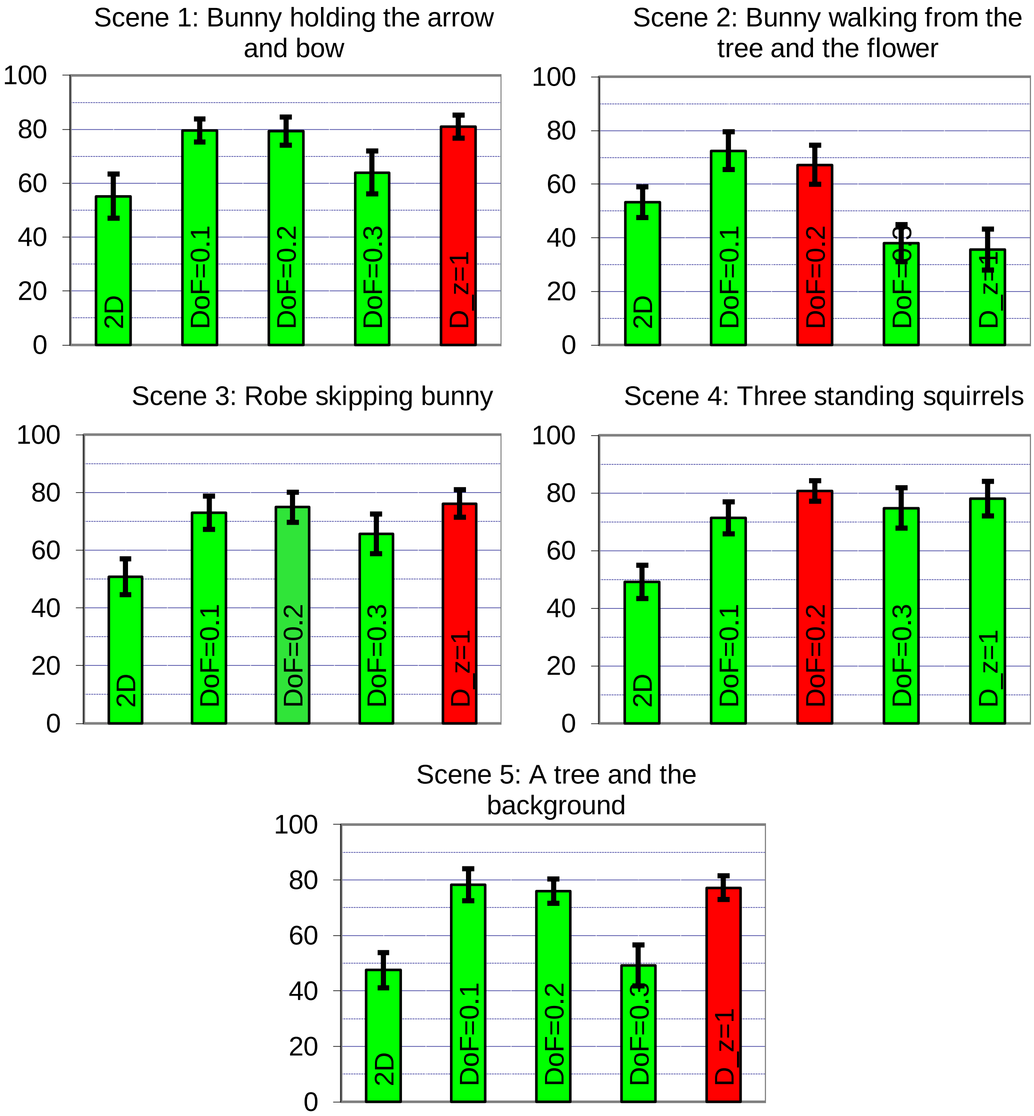


Figure 5-11 : Mean opinion scores and confidence intervals of visual experiences in five different conditions for the different scenes (red/deep color bar is the selected conditions which fulfill the proposed shooting rules in each scene)

Statistical analysis

The analysis of variances (ANOVA) was applied in order to understand if the variation of different parameters in the experiment is statistically significant to the subjective result. “p=0.05” was used for rejecting the null hypothesis. Firstly, a two ways ANOVA was applied considering two factors “Camera baseline” and “Scene”. The P-values are presented in the Table 5-6. “Camera baseline” and “Scene” are all significant for all the subjective indicators. However, it is noticed that the P-value of the “Scene” factor in Visual experience is close to the reject threshold and it is less important than the “Camera baseline” factor for visual experience. The same trend can be observed for depth rendering and visual comfort.

Depth of Focus in diopters was used in this experiment to directly represent the perceived depth range. The second ANOVA analysis was aimed to compare the “DOF” and the “Scene” factors for test conditions (2) to (4) as shown in Table 5-7. For depth rendering, the “DOF” and the “Scene” are both insignificant. The rating of depth rendering may be influenced by the visual comfort, i.e. the subjects tended to report bad depth rendering score when they suffered from visual discomfort. Thus, we get rid of the stimuli rated as visual discomfort (visual comfort score is lower than “fair”) and applied again the ANOVA test. The result showed that “DOF” factor is significant (p=0.035) and “Scene” factor is still insignificant (p=0.326).

For visual comfort, “DOF” and “Scene” are both significant which means that perceived depth range and scene setting are all affecting the visual comfort. However, for visual experience, “DOF” is a significant factor while “Scene” is rejected.

Table 5-6 : P-values of two ways ANOVA (“Camera baseline” and “Scene”)

P-value	Camera baseline	Scene
Depth rendering	6.48E-08	0.014983
Visual comfort	0.000218	0.019501
Visual experience	0.00246	0.043085

Table 5-7 : P-values of two ways ANOVA (“DOF” and “Scene”)

P-value	DOF	Scene
Depth rendering	0.236949	0.169927
Visual comfort	0.000192	0.021414
Visual experience	0.014225	0.183257

In order to understand the relationships among the subjective indicator as well as to prove the proposed priority rule (shooting rule 3), we simply regroup the subjective results into “Discomfort free” (MOS value of visual comfort is equal or above 60-“good”) and “Discomfort problem” (MOS value of visual comfort is below 60-“good”) .19 stimuli and 6 stimuli are grouped into “Discomfort free” and “Discomfort problem” group, respectively. The MOS of visual experience in “discomfort free group” is 70 while it is only 52 in “discomfort problem” group. Absolute Pearson product-moment correlation coefficients were calculated among two pairs of subjective indicators: Depth rendering vs. Visual experience, Visual comfort vs. Visual experience. The result is shown in the Table 5-8. In the “Discomfort free” group, depth rendering has much higher relation with visual experience (coef=0.98) compared to visual comfort (coef=0.58). This confirms the proposed shooting Rule 1 of optimizing the shape distortion in order to improve the depth rendering performance. However, in discomfort problem group, Visual comfort is the dominant factor of visual experience (coef=0.99). These findings likely confirm the proposed priority rule – shooting rule 3.

Table 5-8 : Correlation coefficients among three pairs of subjective indicators

Correlation coefficient *coef*	Depth rendering vs. Visual experience	Visual comfort vs. Visual experience
Discomfort free	0.98	0.58
Discomfort problem	0.79	0.99

5.3.4 Discussion and conclusion

There are several results and findings indicating that the proposed shooting rules and associated priorities can ensure an improved visual QoE (depth rendering, visual comfort and visual experience):

1) Concerning the Rule 1, the optimization of shape distortion showed its advantage in the depth rendering assessment. In the "discomfort free" group, the dominant effect of depth rendering on the visual experience also confirmed this point. However, it was still not easy for subjects to distinguish different depth ranges. It may be due to the synthetic contents which cause the difficulties for the subject to compare the displayed objects with the real world experience. Chapter 6 in this thesis will both use natural content and synthetic content to see whether the content types will have an influence on the results.

2) Considering the Rule 2, visual comfort dropped steeply to be below "good" item if DOF was larger than 0.2 diopters. Thus, guaranteeing the perceived depth within the comfortable viewing zone is mandatory. The comfortable viewing zone is suggested to be around 0.2 diopters.

3) The priority of Rule 2 versus Rule 1 was likely confirmed by the findings that in the "discomfort problem" group, visual comfort was the dominant factor of visual experience. Thus, optimization of visual comfort is prior to optimization for shape distortion.

Chapter 6 The impact of variation of perceived binocular depth on the QoE of S-3DTV

Contents

6.1 Introduction

In the previous chapter, we verified the proposed stereoscopic shooting rules. However, two main questions remain:

1) Synthetic contents were used and subjects seemed to have difficulties to judge the depth rendering of synthetic contents. This may be because people were not familiar with the object and scene in synthetic content. Thus, it is important to add natural contents as stimuli. We assume that people should be more sensitive about the depth in natural content.
2) Only three QoE indicators were used. It may be not enough to understand the perceptual impact of stereoscopic images.

Thus, it is important to design a subjective QoE experiment with more QoE indicators to assess both natural content and synthetic contents. Thus, a new experiment is presented in this chapter. In this experiment, we aim to focus on the exploration of how the most important added value - binocular depth variations – affects the QoE of stereoscopic images. Both natural and synthetic scenes are used as stimuli. For each scene, shooting parameters are selected to generate different stimuli representing different levels of final perceived binocular depth. Six QoE indicators including 2D image quality, depth quantity, visual comfort, depth rendering, naturalness and visual experience are used in the subjective QoE assessment to evaluate the impact of binocular depth variation on the QoE of S-3DTV. The acceptability of visual comfort is also measured by a binary scale (“acceptable” or “not acceptable”) to reveal subjects’ acceptance criteria of visual comfort on S-3DTV. Furthermore, the relationship between different QoE indicators will be investigated and it will lead to propose a new QoE model for S-3DTV.

The chapter is organized as follows: Section 6.2 describes how the experiment contents (both synthetic and natural scenes) are designed and captured in order to generate a variation of binocular depth. Section 6.3 and 0 focus on the subjective QoE assessment which reveals how binocular depth variation affects the different aspects of the QoE on stereoscopic images. Section 6.5 models the higher level concept QoE indicators (depth rendering, naturalness and visual experience) as a weighted sum of basic elements (2D image quality, depth quantity and visual comfort). Concluding remarks are provided in the last section.

6.2 Stereoscopic image (synthetic and natural) generation and capture

In this study, the maximum perceived binocular depth range in the scene is also represented as DOF (depth of focus) in the unit of diopters as the previous chapter. All the camera parameters were calculated in order to represent the same final perceived binocular depth range for each scene. DOF equal to 0.2 diopters was proposed as the threshold of visual comfort in previous chapter. Thus, for each scene, three images at three DOF levels (0.1 diopters, 0.2 diopters and 0.3 diopters) are captured and generated by adapting the shooting parameters (camera baseline) in order to represent the binocular depth variation.

Both natural scenes and synthetic scenes were included. The capture of natural scenes used two professional 2D cameras (camera sensor 8.8x6.6 mm^2) and 3D rigs (mirror rig and side by side rig) in a toed-in setting. All the images were post-processed by professional company after capturing in order to avoid image asymmetry problems. The synthetic scene creation was based on the open animation project “big buck bunny”(BlenderFoundation, 2008) and rendered by the Blender software (virtual camera sensor 32x16 mm^2). Three natural scenes and two synthetic scenes were used as shown in Figure 6-1 and all scene parameters are described in Table 6-1.

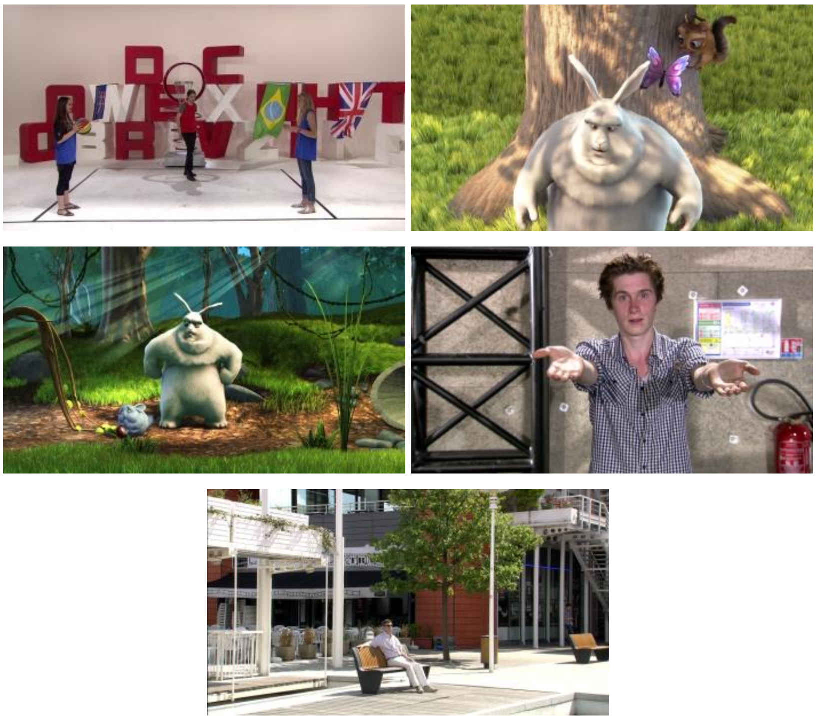


Figure 6-1: Three natural scenes and two synthetic scenes (Top left: Basket; top right butterfly; mid left: Forest; mid right: Interview; bottom: Bench)

The final visualization environment was the same as experiment presented in the previous chapter except the viewing distance. The comfortable viewing zone represented as DOF is also a function of viewing distance. Increasing the viewing distance will increase the comfortable viewing zone. It will provide a larger depth budget and facilitate the optimization of shape distortion. Thus, in this experiment, the viewing distance is 4.5 times of display height compared with 3 times of display height used in Chapter 5.

Shooting parameters were calculated to acquire the perceived binocular depth to guarantee DOF values as 0.1, 0.2 and 0.3 diopters in the final visualization. Table 6-2 depicts these parameters.

Finally, the stereoscopic shape distortion factors, representing the shape distortion around the region of interest (a value of 1.0 indicates no shape distortion, less than 1 means compression in depth, larger than 1 means stretching in depth), are shown in Table 6-3.

Table 6-1 : Scene parameters

Scene Name*	Near (m)	Far (m)	ROI* (m)	Conv* (m)
Basket(N)	5	10	7	5
Butterfly(S)	5.8	12	6.8	6.8
Forest(S)	5	23	7.5	5
Interview(N)	2.6	5	3	2.6
Bench(N)	<14	32	20	14

**N as Natural, S as Synthetic, ROI as Region of Interest, Conv as Convergence*

Table 6-2 : Shooting parameters

Scene Name	Focal (mm)	Camera baseline(mm)		
		DOF 0.1	DOF 0.2	DOF 0.3
Basket(N)	9	160	324	485
Butterfly(S)	70	118	236	353
Forest(S)	36	93	185	278
Interview(N)	22.5	35	65	105
Bench(N)	20	180	362	540

Table 6-3 : Stereoscopic shape distortion

Scene Name	Stereoscopic shape distortion factor		
	DOF 0.1	DOF 0.2	DOF 0.3
Basket(N)	1	2.54	4.76
Butterfly(S)	0.69	1.38	2
Forest(S)	0.55	1.26	2.20
Interview(N)	0.5	1	1.78
Bench(N)	0.41	1.0	1.8

6.3 Experimental setup

A subjective QoE assessment experiment was designed to investigate the effect of different scenes (5 scenes as shown in Table 6-1) and different perceived binocular depth levels on six QoE indicators consisting of 2D image quality, depth quantity, depth rendering, visual experience, depth rendering and visual comfort. The method of this experiment is presented as follows:

1) **Stimuli:** the image materials used in this experiment consisted of three natural scenes and two synthetic scenes are as shown in Figure 6-1. For each scene, there were four images representing the final perceived depth as DOF 0, 0.1, 0.2 and 0.3 diopters respectively. The left view of the stereoscopic image representing 0.1 diopters DOF was used as a 2D image, also referred to as 0 diopter image. 4×5 (DOF $\times$ scene) images were presented in each test session.

2) **Equipment:** the test room and the display were the same as the subjective experiment in Chapter 5. However, the viewing distance was adapted to 2.6 meter as 4.5 times of display height. As what we had shown in Section 3.3.3, increasing the viewing distance will increase the depth rendering ability of the visualization. Theoretically, it should provide a better depth sensation to the viewer.

3) **Observers:** 28 observers were recruited to participate in this test. All of them were non experts in the audiovisual and video domain. The same vision test as Experiment 2 was conducted for each subject. All observers had a normal or corrected to normal visual acuity and normal stereoacuity.

4) **Procedure:** the test consisted of six sessions corresponding to six 3D QoE indicators. Moreover, the subjects were required to report whether they would accept the sequence in terms of visual comfort. Thus, in the visual comfort test session, both 5-levels continuous quality for visual comfort and a binary scale (“acceptable” or “not acceptable”) for acceptability were rated by the subjects. In order to avoid interaction between QoE indicators as well as to avoid accumulating visual discomfort, the whole test was separated into two parts which were conducted on two different days. The first part was composed of three sessions: 2D image quality, depth rendering and visual comfort. The second part also consisted of three sessions: visual experience, naturalness and depth quantity. For each session, there were 4×5 (DOF $\times$ scene) still images presented to viewers for rating. The 20 stimuli were individually randomized for each test session. SAMVIQ method was used to evaluate subject’s opinion of each stimulus on each QoE indicator.

6.4 Result analysis

Figure 6-2 depicts the MOS (mean opinion score) with their 95% confidence intervals per QoE indicator as a function of DOF (increasing along the x-axis) for each scene. A one-way ANOVA analysis was performed with DOF as independent variable and MOS per QoE indicator as dependent variable. The statistical analysis results showed that image quality ($F = 0.96$, $p < 0.436$) was not affected by the variation of binocular depth. The result of depth quantity ($F = 1659$, $p < 0.001$) indicated that the subject can easily distinguish different perceived depth range. And with the increase of perceived depth, visual comfort ($F = 13.30$, $p < 0.001$), decreases significantly as shown in Figure 6-2. Depth rendering ($F = 35.57$, $p < 0.001$), Naturalness ($F = 7.10$, $p < 0.004$) and Visual experience ($F = 9.49$, $p < 0.002$) are similarly affected by the binocular depth variation. When increasing the perceived depth, at the beginning 3D shows advantages over 2D image, e.g. DOF 0 (as 2D) is rated as “poor” in depth rendering, and “fair” in naturalness and visual experience while in DOF 0.1 condition all of these indicators are scored between “good” and “excellent”. However, when the perceived depth is higher than a certain value (DOF 0.2 for Butterfly and Forest, DOF 0.1 for the other scenes), these advantages seem to be reduced. The feedback and discussion with the viewers confirmed that visual comfort should be the main concern which reduced the advantage of added depth.

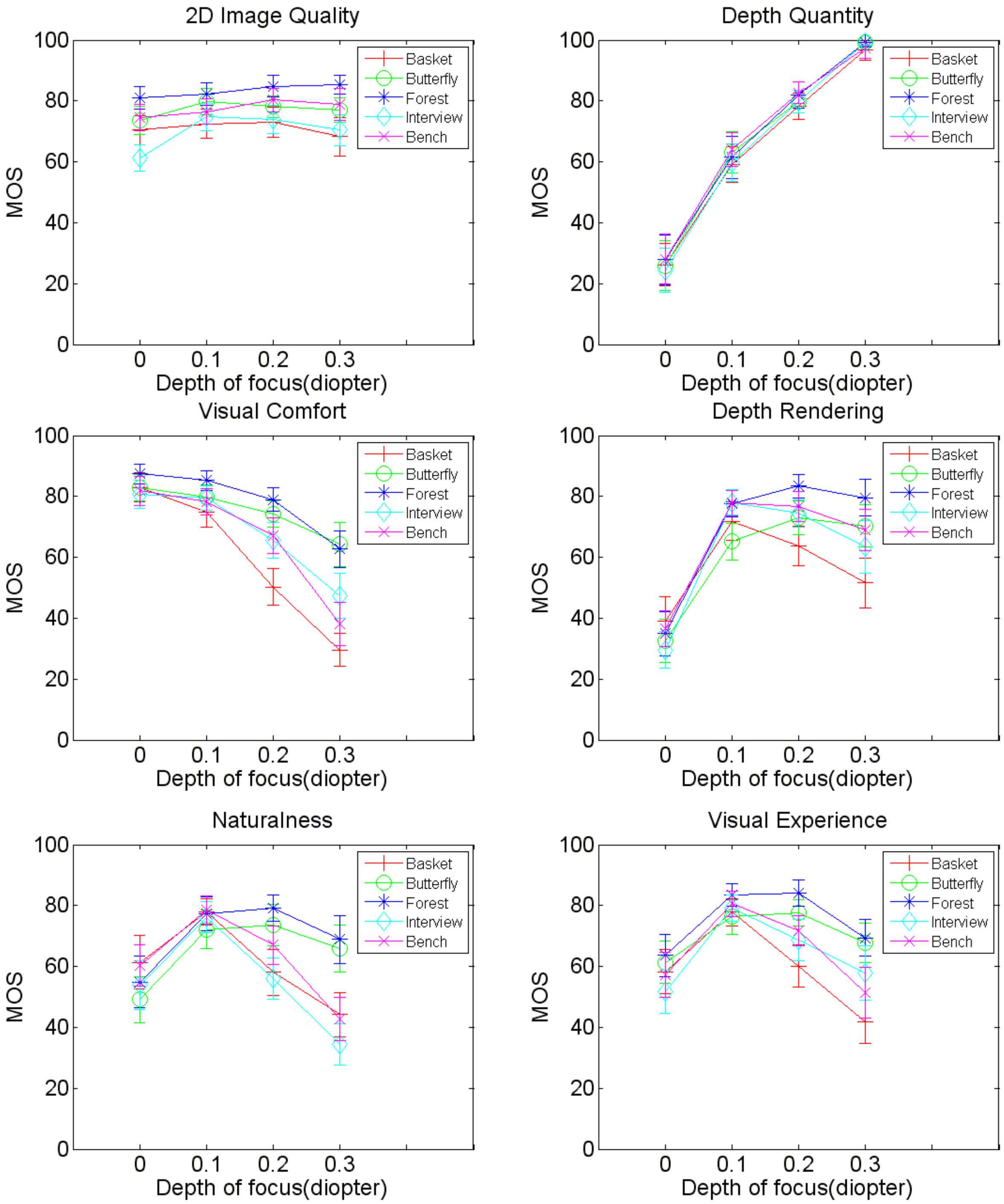


Figure 6-2 : MOS (with their 95% confidence intervals) vs. Variation of DOF for different QoE indicators for different scenes (Basket, Butterfly, Forest, Interview, and Bench as shown in Figure 6-1)

If we consider the shape distortion factor as shown in

Table 6-3, the basket scene in DOF 0.1 and the other scenes in DOF 0.2 should show advantages compared to in other perceived depth conditions, especially in depth rendering. However, there are no significant evidences shown in Figure 6-2, although in Basket (DOF 0.1), Butterfly (DOF 0.2), and Forest (DOF 0.2) conditions, the scores of depth rendering are rated slightly better than the other conditions. This may be due to several reasons, e.g., people are used to viewing 2D images and they are not sensitive to shape distortion in 3DTV especially in the case when the visual discomfort problem is essential.

Figure 6-3 depicts the MOS with their 95% confidence intervals per QoE indicator as a function of DOF between the natural scenes and synthetic scenes.

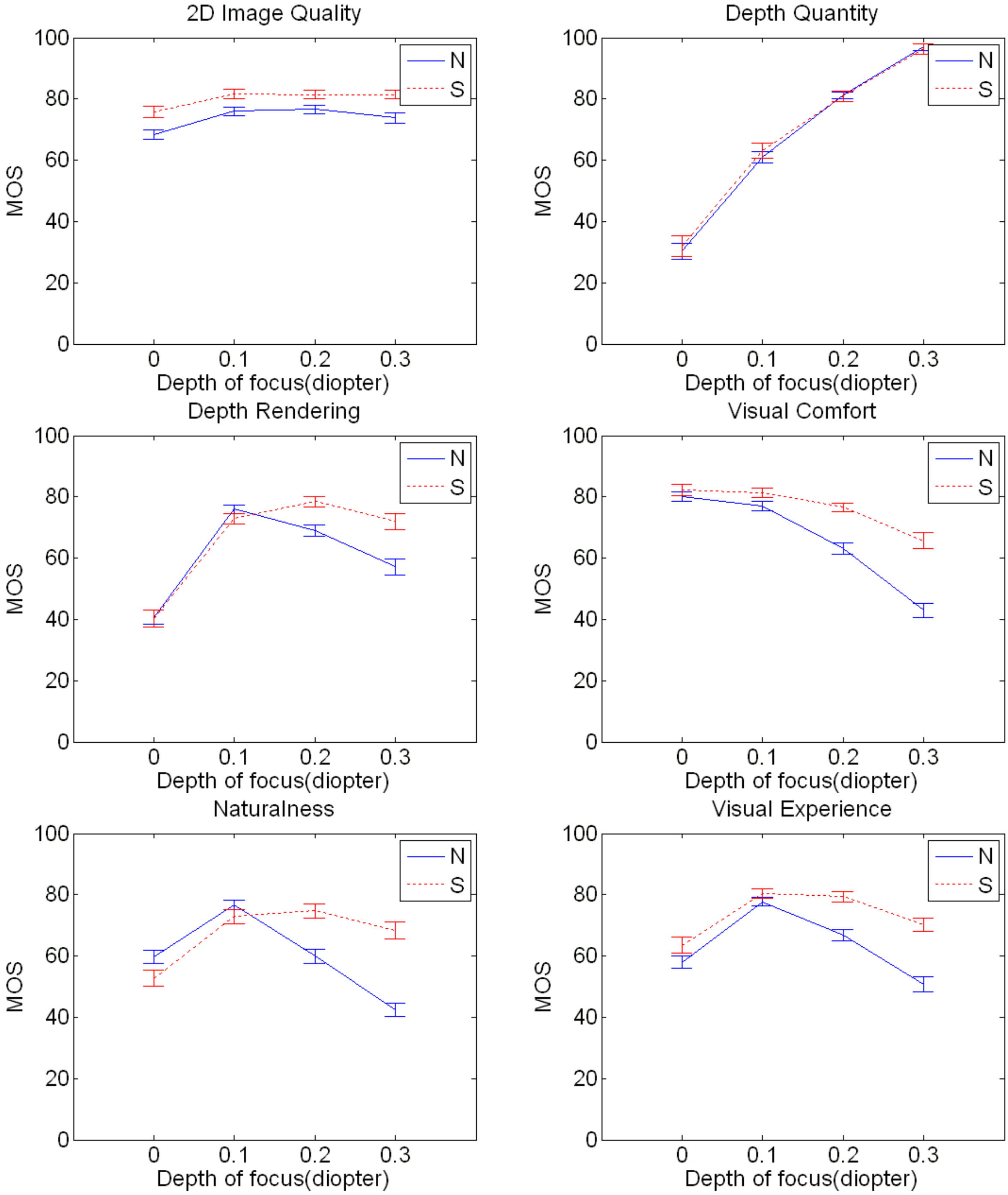


Figure 6-3 : MOS (with their 95% confidence intervals) vs. Variation of DOF for different QoE indicators (Natural scene in solid line and Synthetic scene in dotted line)

In terms of depth quantity and 2D image quality, both synthetic and natural scenes behave similarly. For visual comfort, natural scenes decrease faster than synthetic scenes with the increase of DOF, e.g., in DOF 0.3, synthetic scenes still maintain “good” while natural scenes drop to some value between “fair” and “bad”. There are several possible explanations: firstly, human are used to viewing natural scene compared with synthetic scene; secondly, for natural shooting there exists some performance constraints such as optic focal length, thus blur effect cannot be avoided. For example, the background wall of the “interview” scene is strongly blurred and this blur may cause depth cue contradiction resulting in visual discomfort when people try to focus on the background. For synthetic scenes, all the contents were generated in a way that there appears no blur produced by the focal length and all depth layers are sharp. The same trends between the natural scenes and synthetic scenes are shown in depth rendering, naturalness and visual experience, which may be due to the interaction with visual comfort.

Figure 6-4 depicts the approximated curve of acceptability in different quality grades of visual comfort. The approximation was using MATLAB line fitting function “shape-preserving interpolant”. The results reveal that around 80 percent of subjects accepts the score 60, i.e., between “good” and “fair” on the visual comfort criteria. Only 50 percent of subjects can accept 50, i.e., “fair”. 80 percent are generally used as a rule-of-thumb threshold in many service-oriented applications. Thus, the visual comfort should be maintained as higher than 60. The above finding results in a recommendation for optimized perceived depth: For natural scenes, DOF 0.1 should be targeted and for synthetic scenes, the DOF threshold may remain 0.2.

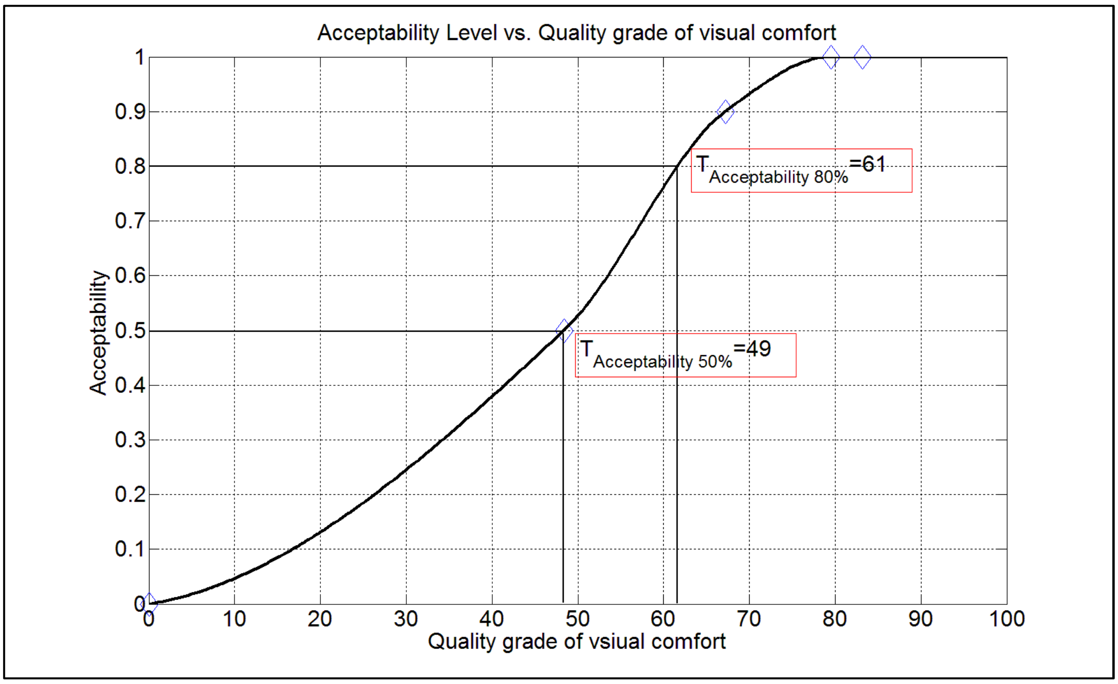


Figure 6-4 : Acceptability vs. Quality grade of visual comfort

6.5 3D QoE modeling

As explained in the previous section, 2D image quality is independent of depth variation while depth quantity and visual comfort shows nearly linear relation with perceived binocular depth. Viewers can judge these three QoE indicators independently so that these three indicators may be categorized as the basic level of

3D QoE aspects. Furthermore, visual experience, naturalness and depth rendering may be defined as higher level of 3D QoE as people need to incorporate the basic level QoE concept in order to form the final perceptual opinion.

A 3D QoE model is proposed in Figure 6-5.

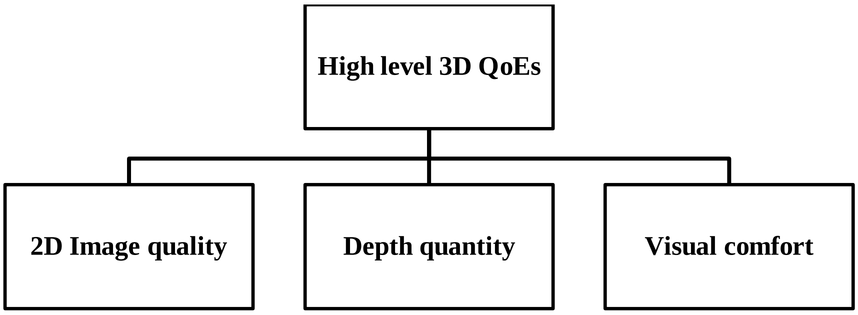


Figure 6-5 : 3D QoE model

Similar to (Lambooij et al., 2011), in order to explore the relationship between the higher level concept and the basic QoE aspect in 3D QoE, we assume that higher level 3D QoE indicators (QoE) can be represented as a weighted sum of 2D image quality (IQ), depth quantity (D) and visual comfort (VC):

$$QoE = \alpha \cdot IQ + \beta \cdot D + \gamma \cdot VC$$

with α, β, γ representing the weights of 2D image quality, depth quantity and visual comfort respectively.

It should be noted that the current purpose of this experiment is less relevant to modeling the 3D QoE by using physical parameters. Instead, the main purpose is to explore in which way high level 3D QoE is formed by basic level concepts. A simple linear regression analysis was performed using the data from this experiment and the coefficients of each component for visual experience, naturalness and depth rendering are shown in the Table 6-4.

Table 6-4 : Weighted coefficients

		IQ	D	VC	R square
Visual experience	Regression	0.205	0.177	0.568	0.973
Naturalness	Regression	0.202	0.137	0.541	0.955
Depth rendering	Regression	0.151	0.366	0.384	0.957

The linear fitting is sufficient to explore the relationship between the higher level QoE concept and the basic level QoE aspect as can be seen by the correlation coefficients (R square > 0.95). The fitted coefficients show that depth quantity influences more on depth rendering (36.6%) than on visual experience (17.7%) and naturalness (13.7%). This also fits for the definition of depth rendering that required viewers to concentrate on the depth and space itself. Visual experience and naturalness scores are determined more by visual comfort (56.8% and 54.1% respectively) than by depth quantity. This also confirmed the proposed shooting rule defined in Section 5.2 that visual comfort is prior to perceived depth in order to guarantee a high overall QoE.

6.6 Conclusion and recommendation

In this experiment, we explored how binocular depth affects the quality of experience of stereoscopic images. The findings are summarized below:

- Increasing the binocular depth does increase the perceived depth quantity as people can easily judge different perceived binocular depth levels. However, at the same time it decreases the visual comfort.
- 2D image quality is not affected by the variation of binocular depth.
- The higher level QoE indicators, depth rendering, naturalness and visual experience may be predicted by a weighted sum of 2D image quality, depth quantity and visual comfort when only variation of binocular depth is considered. The coefficient of linear fitting showed that visual comfort is the dominant factor for visual experience (56.8%) and naturalness (54.1%). This also confirmed the proposed shooting rule 3 defined in last chapter that visual comfort is prior to perceived depth in order to guarantee a high overall QoE.

Moreover, recommendations concerning content production are proposed based on the results:

- For synthetic content, maximum 0.2 diopters of DOF should be targeted to maintain visual comfort.
- For natural content, maximum 0.1 diopters of DOF should be targeted.

Part III Impact of compression, image representation format and view asymmetry on S-3DTV QoE

Chapter 7 The impact of JPEG 2000 compression on the QoE of S-3DTV

Contents

7.1 Introduction

Image compression technique is used to reduce irrelevance and redundancy of the image data in order to store or transmit data in an efficient way. Understanding the impact of image compression technique on the QoE of images will facilitate the selection of optimum compression technique and transmission bitrate for dedicated applications. As introduced in Chapter 1, current S-3DTV broadcast tends to re-use conventional compression technique to compress and transmit stereo image signals.

The impact of conventional image compression techniques on the quality of 2D image is well studied. However, it might not be able to be applicable to S-3DTV application. There are two main reasons: first, the impact of new artifacts such as binocular artifacts (Atanas Boev, 2012) induced by compression of 3D contents on the image quality requires further investigation; second, the QoE of S-3DTV is multi-dimensional including not only 2D image quality but also other QoE indicators such as depth quantity and visual comfort. The impact of image compression techniques on the QoE of S-3DTV is still unknown.

Thus, in this chapter, we aim to investigate the impact of image compression on the QoE of stereoscopic still images. JPEG-2000 compression scheme is used as the compression scheme. The reference 3D scenes including two natural scenes and two synthetic scenes in this study are selected from experiments presented in Chapter 6 in order to avoid visual comfort from image production. Moreover, the left view of each scene is used to represent the 2D image. Both 2D image and each view of stereoscopic 3D images are coded using five different levels of JPEG 2000 compression ratios. A subjective QoE experiment using five QoE indicators (2D image quality, depth quantity, visual comfort, depth rendering and visual experience) is designed to evaluate the impact of JPEG 2000 compression on the QoE of 2D image and stereoscopic 3D images in case of S-3DTV. The result analysis of this study reveals this impact. Furthermore, similar hypothesis as described in Section 6.5 that high level QoE indicators can be estimated by basic level QoE indicators is also reevaluated in this study.

This chapter is organized as: Section 0 presents the experiment design; Section 0 focuses on the analysis of the results from subjective QoE experiment; Section 7.4 present models QoE of S-3DTV; the final conclusion and recommendation are presented in Section 7.5.

7.2 Experimental setup

The experiment design is similar to the experiment presented in Chapter 6. Due to the fact that subjects reported difficulties to evaluate the naturalness on the synthetic contents as well as to reduce the complexity of the subjective experiment, the naturalness concept was not evaluated in this study. The experiment was targeted to investigate the effect of Scene (4 scenes), Dimension (2D, 3D), JPEG 2000 compression ratio (1, 50, 100, 175, 250) on five different QoE indicators consisting of image quality, depth quantity, visual comfort, depth rendering and visual experience. The method of this experiment is presented as follows:

1) **Stimuli:** the image materials used in this experiment consisted of two natural scenes (Bench and Interview) and two synthetic scenes (Butterfly and Forest) as shown in Figure 7-1.

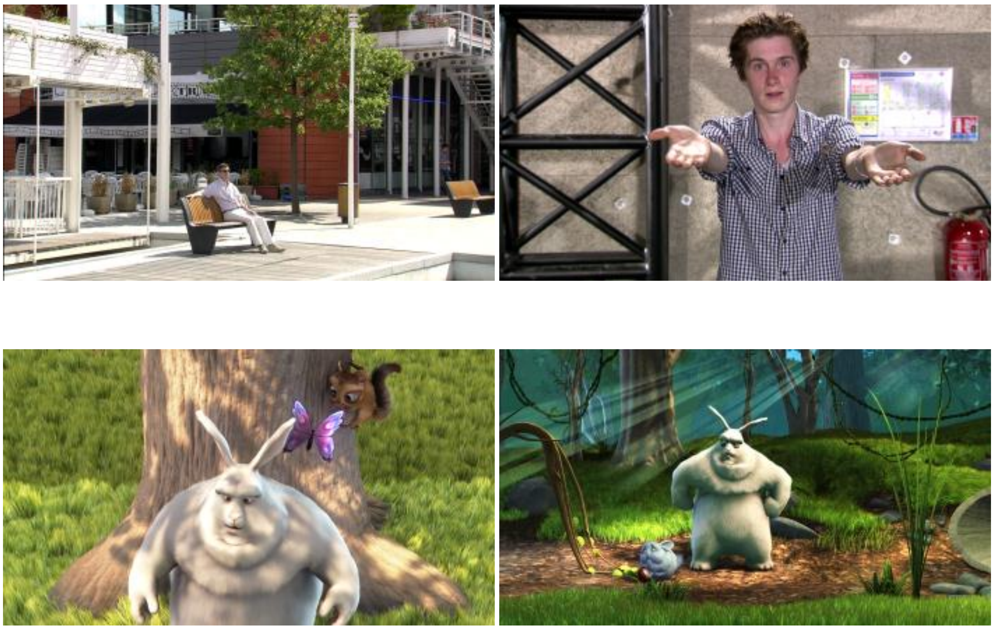


Figure 7-1 : Test scenes (top) left: Bench, right: Interview (bottom) left: Butterfly, right: Forest

The undistorted 3D image (reference) was selected from the experiment presented in Chapter 6 based on two criteria: first, the visual comfort of the stimulus was scored higher than "good" to avoid visual discomfort (93% of subjects accept "good"-70 in MOS of visual comfort as shown in Figure 6-4); second, visual experience of the stimulus was scored as highest in each scene. Thus, for natural scenes, the DOF 0.1 stimuli were selected while for synthetic scenes, the DOF 0.2 stimuli were selected as the undistorted 3D images. The left view of undistorted stereoscopic images was used as the undistorted 2D images. The Jasper JPEG2000 compression software (Adams, n.d.) was used to compress the images. Five compression ratios were selected as 1, 50, 100, 175 and 250 and implemented by "rate" parameter in Jasper software. Figure 7-2 shows the panel images of the 1 and 250 JPEG 2000 compression ratios on the interview scene.

Figure 7-2 : The panel images of the 1(left) and 250(right) of JPEG compression ratios on the interview scene

For 2D images, stimuli were generated by compressing the undistorted 2D images into these five different compression ratios. For 3D images, there were two steps for generating the stimuli: first, the compressed left and right images were generated by compressing the left and right view of 3D images separately; second, since a line-interleaved S-3DTV display was used as visualization terminal, format conversion process were used to convert the compressed full resolution left and right view images to interleaved format. Overall 48 ([Compression level × 2 + 3D explicit reference + 3D hidden reference] ×scene) still pictures were presented in each test session.

2) **Equipment**: the test environment and equipment were identical to the experiment presented in Chapter 6. The viewing distance was fixed to 2.6 meter as 4.5 times of display height.

3) **Observers:** 30 observers were recruited to participate in this test. All of them were non experts in the audiovisual and video domain. The same vision test as the experiment presented in Chapter 6 was conducted for each subject. All observers had a normal or corrected to normal visual acuity and normal stereoacuity.

4) **Procedure:** The test consisted of five sessions corresponding to five QoE indicators including 2D image quality, depth quantity, visual comfort, depth rendering and visual experience. In order to avoid interaction between QoE indicators and in order to avoid accumulating visual discomfort, the whole test was separated into two parts which were conducted on two different days. The first part composed of three sessions: 2D image quality, depth rendering and visual comfort. The second part also consisted of two sessions: depth quantity and visual experience. Written instructions detailing the task were given to the subjects before the start of the test. These instructions were then reiterated by the experimenter as to ensure the observer understood the task. SAMVIQ method was used to evaluate observer's opinion on each stimulus. For each scene, 2D and 3D stimuli were mixed. The explicit reference and hidden reference were the undistorted 3D image of each scene. Thus, the viewer was required to rate 12 images per scene and 48 images per session.

7.3 Result analysis

2D Image quality

Figure 7-3 shows the mean opinion score (with 95% confidence intervals) of 2D image quality averaged over all scenes. On the x-axis the different JPEG 2000 compression ratios are presented (increasing compression ratio along the x-axis). The y-axis presents the MOS of 2D image quality from bad to excellent as numerical scale from 0 to 100. The two lines in the figure represent 2D as solid line and 3D as dashed line. Error bars reflect the standard error of the mean (95% confidence interval).

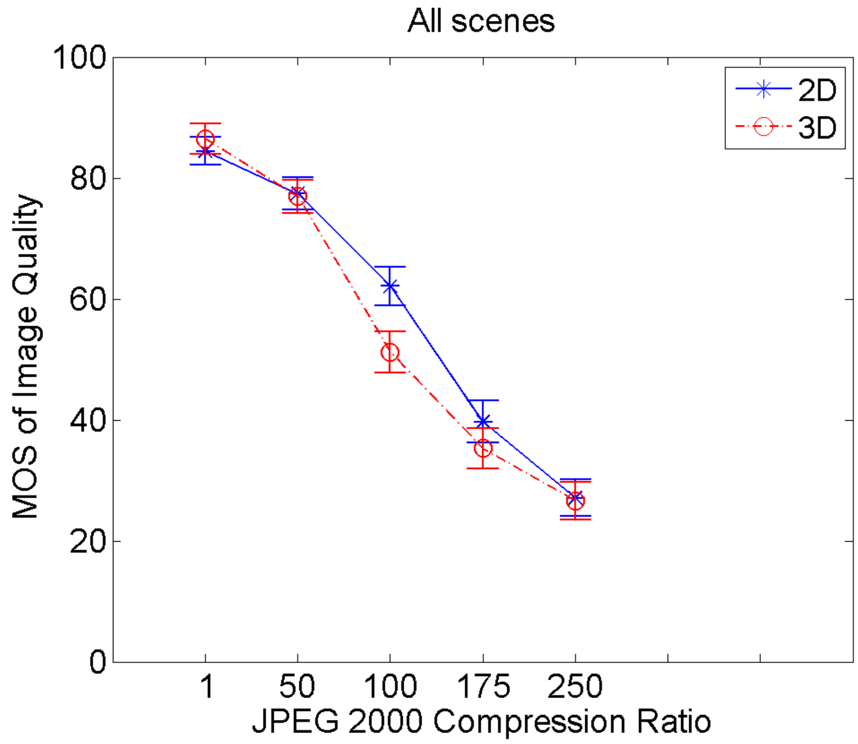


Figure 7-3 : MOS of 2D image quality averaged over all scenes (with their 95% confidence intervals) vs. Variation of JPEG 2000 compression ratio for 2D (solid line) and 3D (dash line) conditions

A one-way ANOVA (with Scene, Dimension and compression ratio, respectively) was carried out on the raw subjective ratings to test the main effects. The result revealed significant main effects of compression ratio ($p<0.001$) and scene ($p<0.019$). Dimension ($p<0.272$) is not a significant factor. The main effect of scene was mainly caused by different texture complexity. The main effect of compression ratio was clearly visible in all images. Figure 7-3 clearly shows the main effect of a decreasing image quality with increasing JPEG 2000 compression ratio for both 2D and 3D images. It also shows that the image quality of 3D image reduces faster than 2D image with the increase of the compression level. Without compression or only low compression ratio such as 50, the MOS with 95% confidence interval of 2D and 3D are similar, overlapping each over. When compression ratio increases to 100, the image quality of 2D images is rated as more than 60 (close to “good”) MOS higher than 3D image as around 50 (close to “fair”). When compression ratio is higher than 100, the difference in image quality between 2D and 3D image reduces with increasing compression ratio. At a compression ratio 250, both 2D and 3D images were rated the same as around 25 MOS.

Depth quantity

Figure 7-4 shows the mean opinion score (with 95% confidence intervals) of depth quantity averaged over all scenes. The ANOVA analysis results revealed significant main effects of dimension ($p<0.001$), compression ($p<0.001$) and scene ($p<0.023$). The depth quantity of 3D images is rated systematically higher than the depth quantity of 2D images, explaining the main effect of dimension. The depth quantity of both 2D and 3D images reduces with increasing compression ratios, explaining the main effect of compression ratio.

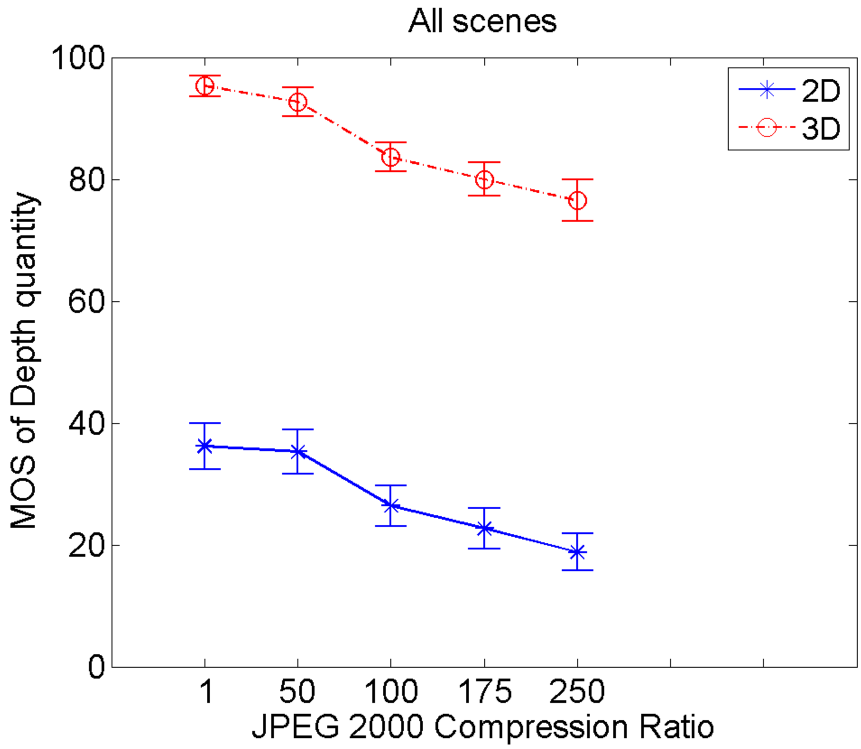


Figure 7-4 : MOS of depth quantity (with their 95% confidence intervals) averaged over all scenes vs. Variation of JPEG 2000 compression ratio for 2D and 3D conditions (2D is solid line and 3D is in dashed line)

The evaluation criteria of depth quantity takes into account both the monocular depth cues and binocular depth cues. 2D images only provided monocular depth cues, but 3D images can provide additional binocular depth cues. The difference in depth quantity between 2D and 3D image in the same compression ratio seems to be constant as around 60 MOS score as shown in Figure 7-4. This indicates that the binocular depth cue greatly enhances the sensation of depth quantity.

Visual comfort

Figure 7-5 shows the mean opinion score (with 95% confidence intervals) of visual comfort averaged over all scenes. The results of ANOVA analysis revealed significant main effects of compression ratios ($p<0.001$) and dimension ($p<0.001$). Scene ($p<0.107$) is not a significant factor for visual comfort. The visual comfort of both 2D and 3D images reduce with increase of compression ratio, explaining the main effect of compression ratio. In uncompressed image or low compression ratio, the visual comfort level of 2D and 3D is similar. However, the difference in visual comfort between 2D and 3D increases with increasing compression ratios as observed in Figure 7-5. Viewers experienced higher level of visual discomfort in 3D image than 2D image with increasing compression distortion. This may be because of more visual artifact induced in 3D images than in 2D images by compression distortion.

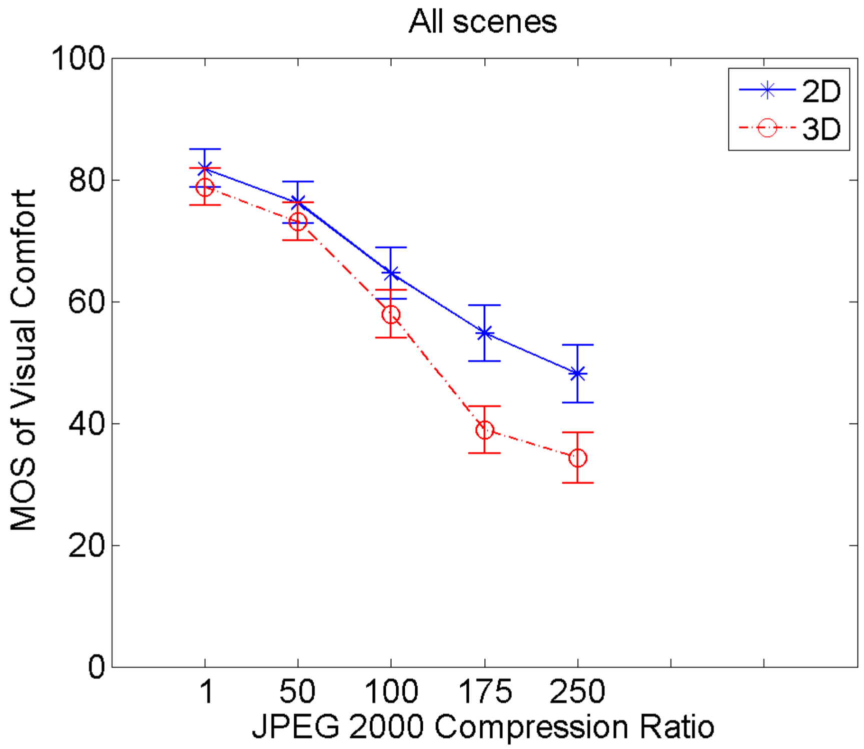


Figure 7-5 : MOS of visual comfort (with their 95% confidence intervals) vs. Variation of JPEG 2000 compression ratio for 2D and 3D conditions (2D is solid line and 3D is in dashed line)

Depth rendering

Figure 7-6 shows the mean opinion score (with 95% confidence intervals) of depth rendering averaged over all scenes. The result of ANOVA analysis revealed significant main effects of dimension ($p<0.001$), compression ratios ($p<0.001$) and scene ($p<0.008$). The depth rendering of 3D images is scored on average 40 MOS higher than the depth rendering of 2D images as shown in Figure 7-6.

Visual experience

Figure 7-7 shows the mean opinion score (with 95% confidence intervals) of visual experience averaged over all scenes. The result of ANOVA analysis revealed significant main effects of compression ratios ($p<0.001$), dimension ($p<0.001$) and scene ($p < 0.003$). 3D images without any compression were rated as “excellent” while 2D images without any compression were only rated as “good”. When compression ratio is smaller than 100, the difference in visual experience between 2D and 3D is equivalent to a change in MOS of around 20. 3D images showed advantages to 2D images in visual experience. At a compression ratio 100, this difference reduces to a MOS difference of around 10. At a compression ratio higher than 100, the advantage of 3D images in terms of visual experience disappeared as the MOS of 3D was rated as the same as 2D.

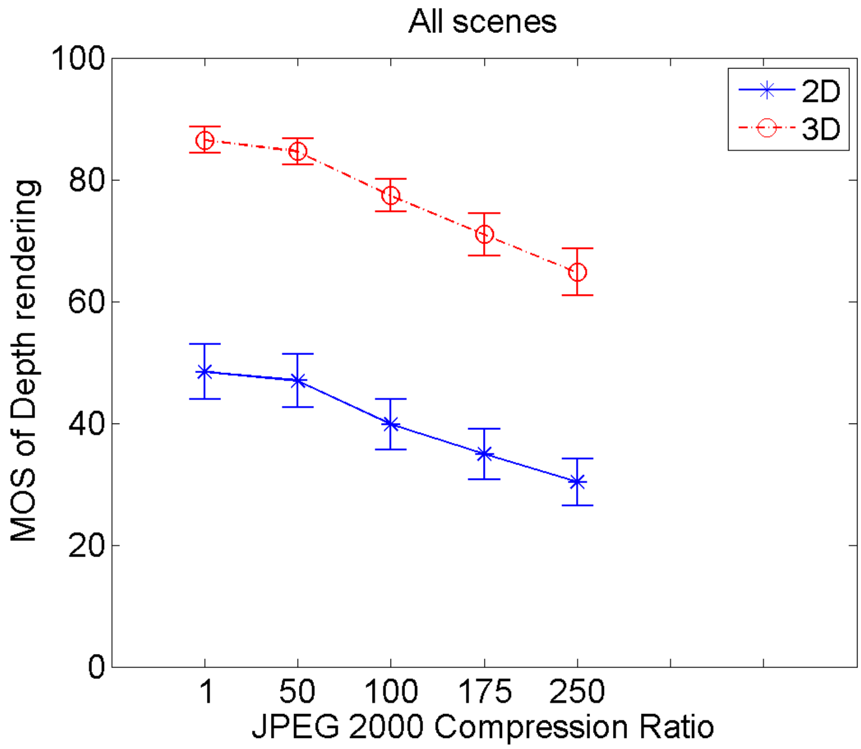


Figure 7-6 : MOS of depth rendering (with their 95% confidence intervals) vs. Variation of JPEG 2000 compression ratio for 2D and 3D conditions (2D is solid line and 3D is in dashed line)

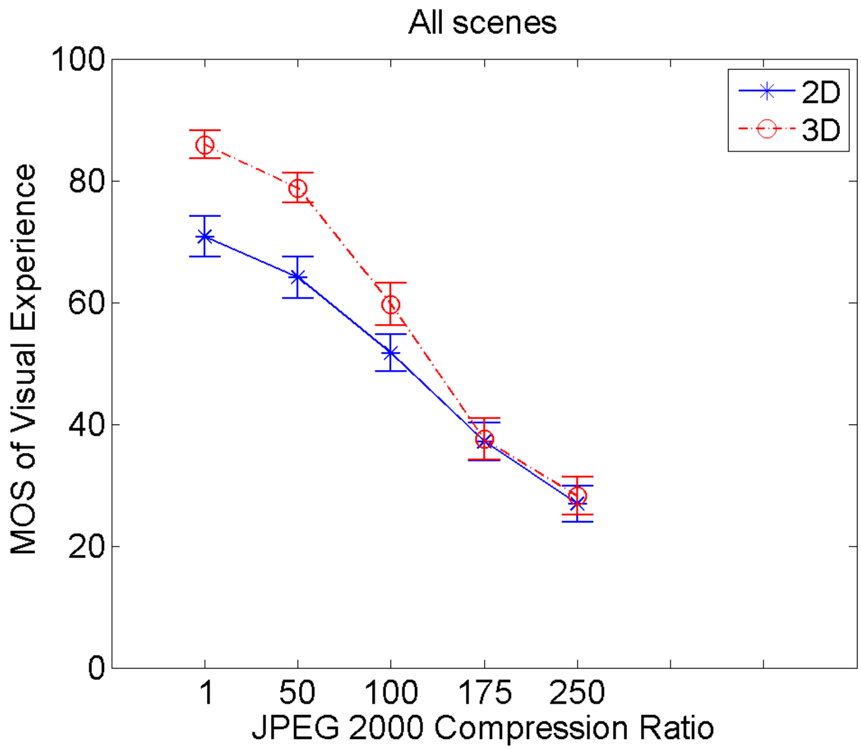
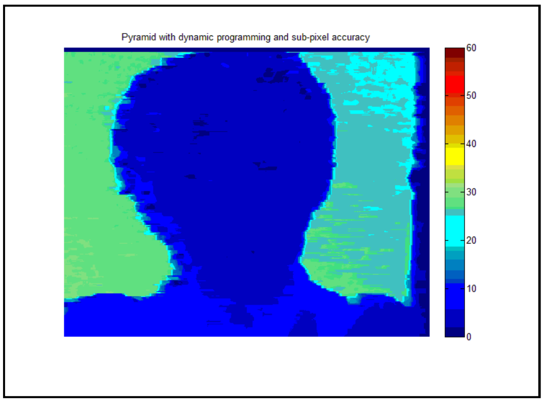
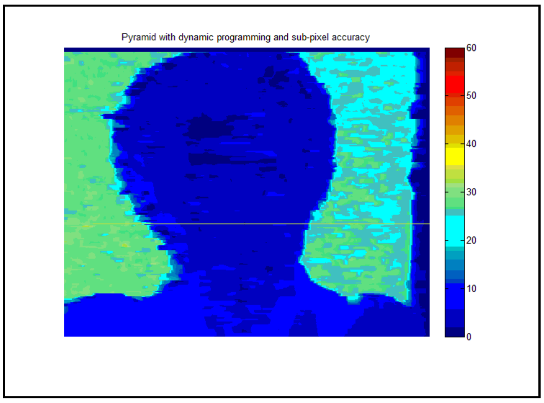


Figure 7-7 : MOS of visual experience (with their 95% confidence intervals) vs. Variation of JPEG 2000 compression ratio for 2D and 3D conditions (2D is solid line and 3D is in dashed line)

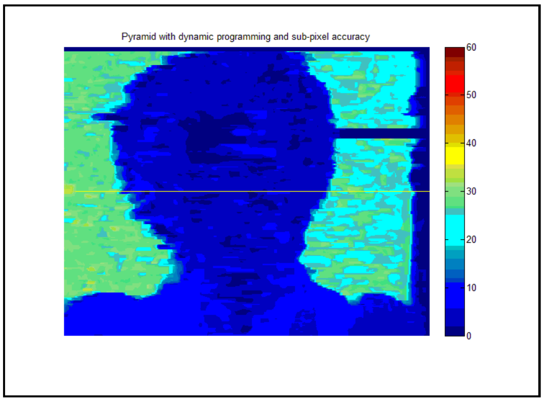
In order to further understand the effect of compression distortion on 3D image, the estimated depth maps of all the 3D images are generated using a robust stereo disparity estimation method (Park and Park, 2001). Figure 7-8 plots the estimated depth map (partial) of the interview scene in different compression levels. More apparent visual artifacts as well as more depth discontinuities can be observed in the depth map with higher compression ratios. It confirms that more visual artifacts are induced in 3D especially in higher compression ratio. This might explain why image quality of 3D reduces faster than 2D and why more visual discomfort is experienced in 3D images with increasing compression distortion.



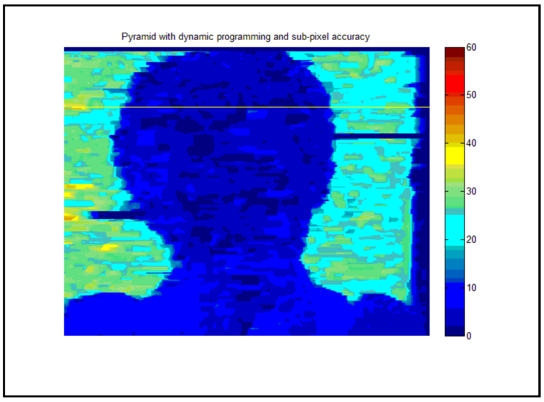
Compression Ratio 1



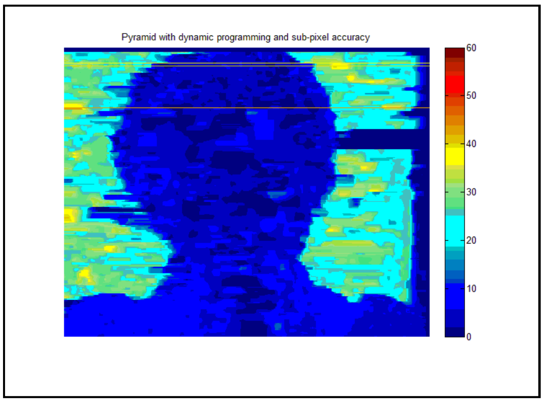
Compression Ratio 50



Compression Ratio 100



Compression Ratio 175



Compression Ratio 250

Figure 7-8 : Estimated depth map in five different JPEG compression ratios

7.4 3D QoE modeling

As introduced in Section 6.5, higher level QoE indicator (e.g., visual experience and depth rendering) can be represented as a weighted sum of basic level QoE indictors (2D image quality, depth rendering and visual comfort). A linear regression is applied to raw subjective ratings and the weighted coefficients are shown in Table 7-1.

Table 7-1 : Weighted coefficient

		IQ	D	VC	R square
Visual experience	Regression	0.507	0.199	0.248	0.939
Depth rendering	Regression	0.205	0.541	0.204	0.888

The linear fitting is sufficient to explore the relationship between the higher level QoE concept and the basic level quality aspect as can be seen by the correlation coefficients (R square > 0.88). The regression results show that visual experience can be determined by a weighted sum of 50.7% of image quality, 19.9% of depth quantity and 24.8% of visual comfort. The depth rendering can be predicted by a weighted sum of 20.5 % of image quality, 54.1% of depth quantity and 20.4 % percent of visual comfort in this study.

7.5 Conclusion and recommendation

The main findings from this study can be summarized as:

1) The JPEG 2000 compression had a global degradation on all the QoE indicators. Increasing the compression ratio, the MOS reduced significantly.
2) Comparing the effect of compression distortion between 2D and 3D images, visual experience of 3D images showed advantages to 2D images at low compression ratios (i.e., less than 100). It can be explained by the added binocular depth. However, this advantage kept reducing with the increase of compression ratios which may be related to more visual artifacts added in 3D by the compression distortion. At higher compression ratios, both image quality and visual comfort were rated lower in 3D than in 2D.
3) The advantage of depth quantity between 2D and 3D did not reduce seriously even in high compression ratios. It may indicate that JPEG compression did not destroy the information of binocular depth cue.
4) The linear regression results (performed on MOS) showed that visual experience in this study can be predicted by a weighted sum of 50.7% of image quality, 19.9% of depth quantity and 24.8% of visual comfort. Moreover, depth rendering can be predicted by a weighted sum of 20.5 % of image quality, 54.1% of depth quantity and 20.4 % percent of visual comfort in this study.

The main recommendation from this study is that at high compression ratio or low bitrate application scenario, there is no interest to provide 3D service because in this scenario: 1) 3D images are not able to provide higher level of visual experience than 2D images; 2) Lower level of quality and visual comfort might be perceived in 3D images than in 2D images.

Chapter 8 The impact of image representation formats on the QoE of line interleaved S-3DTV

Contents

8.1 Introduction

One of the very important problems of the 3D broadcasting chain is the selection of S-3D representation format. Two full resolution views are the ideal choice but it will possibly cause unacceptable bitrate increment resulting in the requirement of new standards and equipment for compression and delivery. Hence, different strategies such as half horizontal resolution as Side-by-Side, half vertical resolution as Top-and-Bottom are used in the industry in order to be frame compatible to conventional HDTV broadcast formats. However, their potential effects on the quality of experience of S-3DTV are still unknown. The interaction among video representation formats, video signal scan type (interlaced or progressive) and S-3DTV display techniques (line interleaved, column interleaved or active shutter) may affect the final quality of experience.

In this chapter, we aim to investigate the influence of video representation formats on the perceived quality of experience on line interleaved 3DTV. In this study, video signal sources cover progressive and interlaced content. Test videos are carefully selected based on the 3D video complexity categorization rule (texture, motion, depth). Different S-3D video representation formats with different levels of horizontal, vertical or mixed (both horizontal and vertical) resolution reduction are simulated.

Two experiments were designed: the first one focuses on the direct comparison of different video representation formats without any compression in line interleaved display. A subjective assessment using SAMVIQ and two QoE indicators (visual experience and depth rendering) are used to evaluate the performance of different video formats. The second experiment compared the QoE of Side-by-Side, Top-and-Bottom and 2D HD video under various bitrates. SAMVIQ method with visual experience indicator was used to evaluate the QoE. The results from this study reveal the impact of different video representation formats on the perceived QoE on line interleaved display.

This chapter is organized as: in the introduction part, we introduce the line interleaved S-3DTV, the interlaced and progressive video signals and different 3D stereo video representation formats. Section 8.2 presents the Experiment 1 of this chapter which focuses on evaluation of different resolution reduction effect on line interleaved display. Section 8.3 presents the Experiment 2 in this chapter which focuses on comparison of Side-by-Side, Top-and-Bottom and 2D HD on different bitrates. Section 8.4 draws the conclusion and recommendation of this chapter.

8.1.1 Line Interleaved 3DTV

Line Interleaved 3DTV uses a special polarized filter (linearly or circularly polarized) in front of the display panel, separating spatially the odd lines and the even lines to be left view and right view respectively. Users receive the left view and right view by wearing the polarized glasses. Figure 8-1 shows the principle of line interleaved display. This filter consists of a P1 type filter in the odd line and P2 type filter in the even line (for linearly polarized solution, P1 and P2 filter should be orthogonal; for circularly polarized solution, P1 and P2 filter should be clockwise and counter clockwise, respectively) in order to filter the left and right image respectively. Finally, by wearing a pair of polarized glasses, left and right image can be perceived by the left and right eye separately.

As presented in Chapter 3, due to full horizontal resolution per view, line interleaved display maintains same depth rendering performance as full resolution display. However, its vertical resolution per view is halved. Thus, the effect of 3D representation format with vertical resolution reduction might be more critical in this display.

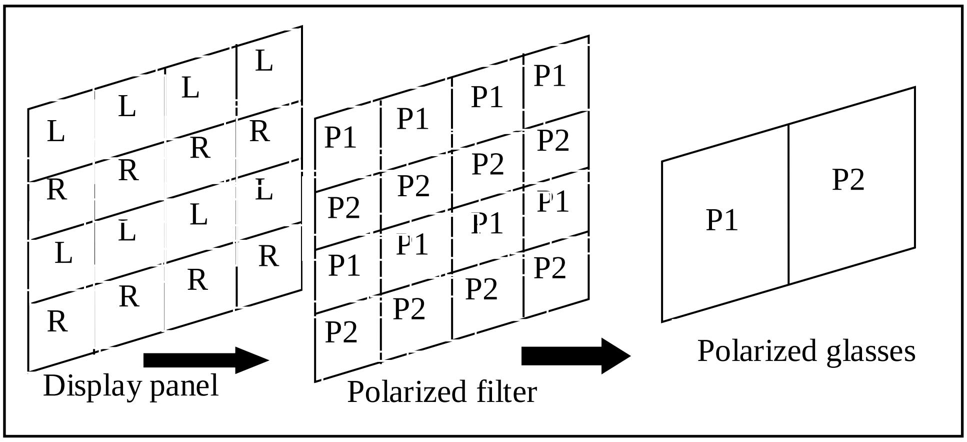


Figure 8-1 : Principle of line interleaved display

8.1.2 Interlaced and progressive video signal

There are two types of video signals available for television signal broadcasting: interlaced video signal and progressive video signal.

For interlaced video signals, a frame contains two fields captured on different time. One field contains all the odd lines of the image and the other contains all the even lines of the image. For progressive video signals, as opposed to interlaced, a frame contains the entire image line by line.

The main interest of interlaced video signals is that in the same bandwidth, it can provide a video signal with twice the temporal resolution versus progressive video signal. For example, 1080i25 provides 50 half-frames per second while 1080p25 provides 25 full frame per second. Mainly cathode ray tube (CRT) display can natively display the interlaced video signal due to the electronic scan. For Liquid Crystal Display (LCD), de-interlacing is required.

Since most of the HD displays are non-CRT but progressively scanned flat panel display, progressive content is recommended by EBU Technical Recommendation R115 (EBU, 2005). However, practically, in professional broadcast domain, interlaced signals are still widely used. For 3DTV broadcasting, most of the television 3D contents still are interlaced video signals. In the opposite, most of the movie contents provide progressive signals. Thus, in this study, both interlaced and progressive video signal will be used.

8.1.3 Different 3D stereo video representation formats

Stereoscopic video should contain left and right eye images so that twice capacities may be required for storage of the uncompressed video. After video compression, transmission bit rate may still be higher than conventional 2D video. In order to be frame compatible with current HDTV formats, normally half horizontal or half vertical resolution technique for left and right eye images is applied. 3D video formats used in this study are categorized as follows:

1) **Half horizontal resolution for each view**, relates to Side-by-Side frame compatible video format as specified in DVB document A154 (DVB, 2011). For example, each view contains 960x1080 pixels and it can be transmitted in the same way as an HD progressive frame of 1920x1080 pixels.

2) **Half vertical resolution for each view,** relates to Top-and-Bottom format Side frame compatible video format as specified in DVB document A154. For example, for a conventional HD frame, it can consist of two views which each view contains 1920x540 pixels.

3) **Reduced both horizontal and vertical resolution for each view**, relates to multi-views (more than 2 views) autostereoscopic display. For example, a 9 views autostereoscopic display normally only has one third horizontal and vertical resolutions for each view which means 640x360 pixels per view for a 1920x1080 pixels panel display.

4) **Full resolution for each view,** i.e., 1920x1080 pixels for each eye in the HD cases.

The schematic diagram of the above four types of 3D video format are illustrated in Figure 8-2.

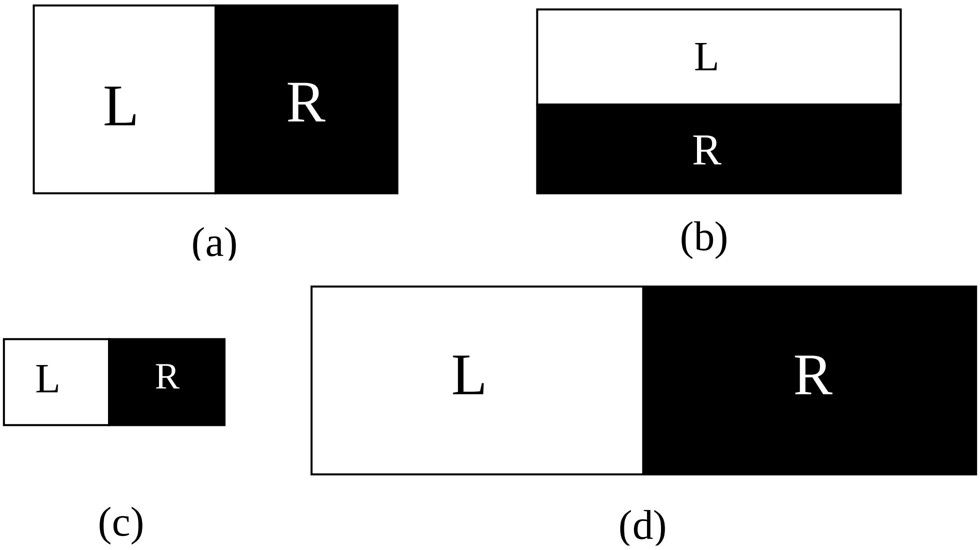


Figure 8-2 : Schematic diagram of four different 3D video format: (a) Side-by-Side with each view of 940x1080 pixels**, (b)** Top-and-Bottom with each view of 1920x540 pixels**, (c)** 1/3 horizontal and 1/3 vertical resolution with each view of 640x360 pixels**, (d)** Full resolution for each view with each view of 1920x1080 pixels

8.2 Experiment 1

The experiment was designed to investigate the effect of different 3D video representation formats (horizontal resolution reduction, vertical resolution reduction, mix resolution reduction and full resolution formats) on two QoE indicators: visual experience and depth rendering.

8.2.1 Methodology

1) Stimuli: Progressive content in 1080p25 and interlaced content in 1080i25 were both integrated in our test stimuli. Progressive stimuli were movie content while interlaced stimuli are television content. We selected the samples or clips based on three video complexities (texture, motion and depth). Considering three levels for each criterion (low, mid and high), 27 combinations of video complexity were possible. However, for constraints of the experiment time and range, it was impossible to cover all these 27 combinations. It was therefore essential to select 6 scenes ranged from a low complexity to high complexity. Six scenes were used in the test as shown in Table 8-1.

For each scene, in order to simulate different 3D representation formats, the following stimuli in different resolution reduction ratios were generated:

- **Horizontal resolution reduction ratios:** 0.66 (2/3 horizontal reduction), 0.5 (Side-by-Side) and 0.375 (3/8 horizontal reduction)
- **Vertical resolution reduction ratios:** 0.66 (2/3 vertical reduction) and 0.5 (Top-and-bottom).
- **Mixed (both horizontal and vertical) resolution reduction:** 0.44 (2/3 horizontal and 2/3 vertical), 0.25 (1/2 horizontal and 1/2 vertical) and 0.11 (1/3 horizontal and 1/3 vertical)

Table 8-1 : Six selected sequences for test

Name	Texture	Motion	Depth	P / I(*)	N / S(*)
Football1	Mid	Mid	High	I	N
Football2	Mid	High	High	I	N
Tennis	High	Low	High	I	N
Tribune	Mid	Low	High	I	N
Ciné1	Low	Low	High	P	S
Ciné2	Low	Low	Low	P	S

(*) P for Progressive and I for Interlaced, N for natural content and S for synthetic content

The cases of "Horizontal 0.5" and "Vertical 0.5" represent respectively the well-known 3D video formats "Side-by-Side" and "Top-and-Bottom". The case of "Mixed (both horizontal and vertical) 0.11" is introduced as an anchor of low quality which related to 9 views autostereoscopic display.

Lanczos3 filter was used to make the effect of resolution reduction. The resolution per view after resolution reduction for progressive content and interlaced content is shown in Table 8-2. It is important to clarity that for interlaced content, one frame contains two fields captured in different times. Thus, view spatial resolution is halved compared to progressive format. However, its temporal resolution is doubled. Moreover, in order to avoid the display internal process, all the samples with resolution reduction were converted into line interleaved format in 1080p50 using internal software (see Appendix B. Representation format conversion). For progressive format, each frame was duplicated once to reach the frame rate 50Hz.

Table 8-2 : The resolution per view under different resolution reduction ratios

Stimuli	P/I*	1	0.66	0.5	0.44	0.375	0.25	0.11
Reference	P	1920x1080						
	I	1920x540						
Horizontal	P		1280x1080	960x1080		720x1080		
	I		1280x540	960x540		720x540		
Vertical	P		1920x720	1920x540				
	I		1920x360	1920x270				
H and V	P				1280x720		960x540	560x360
	I				1280x360		960x260	560x180

P for progressive content, I for interlaced content

To summarize, for each scene, observers were asked to judge 11 video sequences including the reference sequence (undistorted 3D video), the hidden reference sequence (undistorted 3D video), the undistorted 2D hidden reference (left view of undistorted 3D video) and the aforementioned 8 types of reduced resolution sequences.

2) Apparatus and test environment: the stereoscopic video pairs were displayed on Hyundai S465D 46 inches S-3D display which is line interleaved technique based solution (1920x1080 definition panel). Observers wore the circular polarized glasses to watch the images on the S-3D display. A DVS (digital video system) equipment, which was a hard disk raid based digital video record and player capable of playing real-time HD videos up to 1080p 60Hz, was used to output video signals to the S-3D display. A Window-XP based Notebook running the advanced SAMVIQ software was used to control stimuli playback and also rate the quality of experience scores by the tester. The equipment setup is illustrated in Figure 8-3.

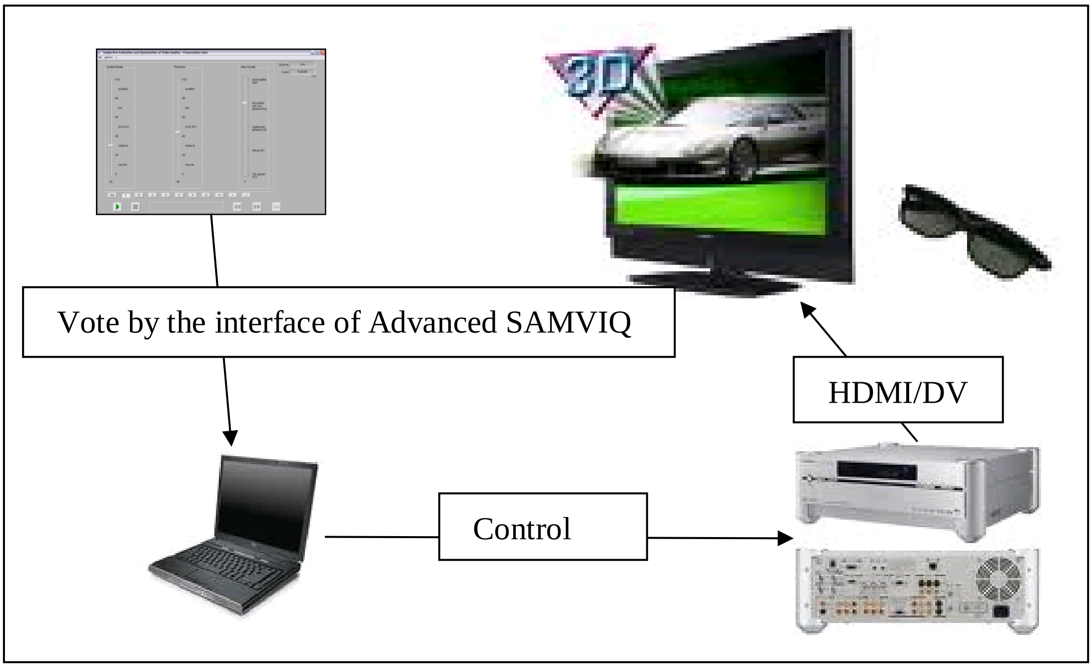


Figure 8-3 : Equipment setup of Experiment 1

To ensure the validity of the results (reliability, reproducibility and etc.), the test was conducted in a lab environment as shown in Figure 8-4 based on the recommendation of ITU-BT.500, including:

- a viewing distance of three times the height of the image
- a maximum luminance of the screen at 100 cd/m^2 through polarized glasses (Note: the maximum luminance of screen measured without glasses is 250 cd/m^2)
- a ratio of 10% between the brightness of background and the peak brightness of the screen

Figure 8-4 : Test environment

3) Observers: Twenty eight observers (18 male and 10 female, age range 20-45) were recruited to participate in this experiment. All observers were non-experts in the audiovisual and video domain. A vision test was performed using Essilor

ERGOVISION equipment on all the testers to determine their visual performance and the potential impact on results. The vision tests showed that all the testers were able to perceive binocular depth, however, with different levels in stereoacuity. Approximately half of the testers (13 of 28 people) could not discriminate a difference of 1 minute of arc. Their stereoacuity was 2 or 3 minute of arc, which was clearly above the detection limit of between 2 seconds to 30 seconds of arc (Patterson, 2007).

4) Procedure: Observers were seated at a viewing distance of 1.8 meter (3 times of picture height) from the stereoscopic display. They were given written instructions detailing the task they had to perform, and the attribute they were asked to rate. These instructions were then reiterated by the experimenter to ensure the observer understand the task at hand.

The stereoscopic videos were presented on the 46 inch line interleaved Hyundai S-3D display, placed in a dimly lit test room following the ITU BT.500 recommendation. There were 6 scenes, each scene consisting of 11 test sequences. The duration of each test sequence was 15 seconds, longer than suggested “less than 10s” in ITU BT.500, in order to facilitate the observer’s immersion of the representative scene. SAMVIQ method was served as the test protocol for this experiment to rate each QoE indicators.

8.2.2 Result analysis

Visual experience

Figure 8-5 shows mean opinion scores of visual experience averaging over all scenes. It decreases with reduced image resolution per view. Only the 3D hidden reference, 2D video and the horizontal reduction limited to 0.5 (as Side-by-Side format) are rated as “excellent”. Moreover, the visual experience of the 3D hidden reference is considered superior to 2D. With the same total number of pixels, horizontal reduction is rated higher than the vertical case. It might be because line interleaved display originally has the effect of vertical resolution reduction (half vertical resolution per view). Further vertical resolution reduction might be more visible than horizontal resolution reduction. It indicates that horizontal reduction is better adapted to the line interleaved display. The 1/9 ratio of horizontal and vertical definition reduction can represent the 9 views autostereoscopic display. Its visual experience is rated apparently lower than traditional stereoscopic display.

Depth rendering

Figure 8-6 shows mean opinion scores of depth rendering averaging over all scenes. Firstly, we can observe that all 3D videos were rated better than 2D. Moreover, the depth rendering rating also decreases with decreasing image resolution. The curves of depth rendering rating have a similar trend as the curves of visual experience rating. At the same reduction ratio of 0.5, horizontal reduction achieved better depth rendering than vertical reduction.

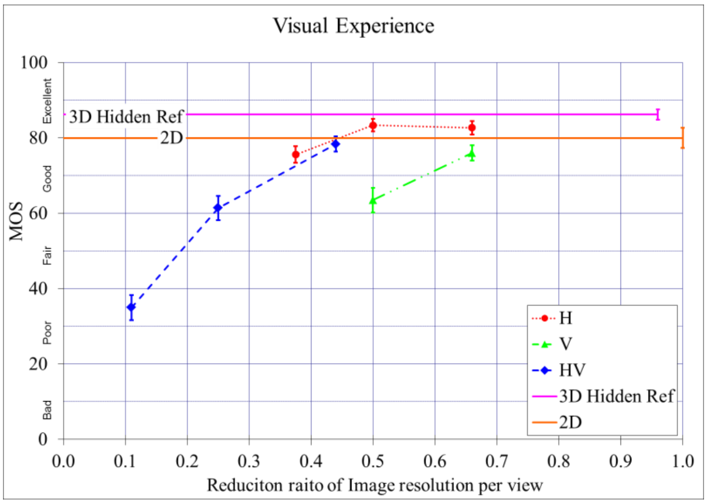


Figure 8-5 : MOS of visual experience averaging over all scenes (with their 95% confidence intervals) vs. Reduction ratio of Image resolution per view

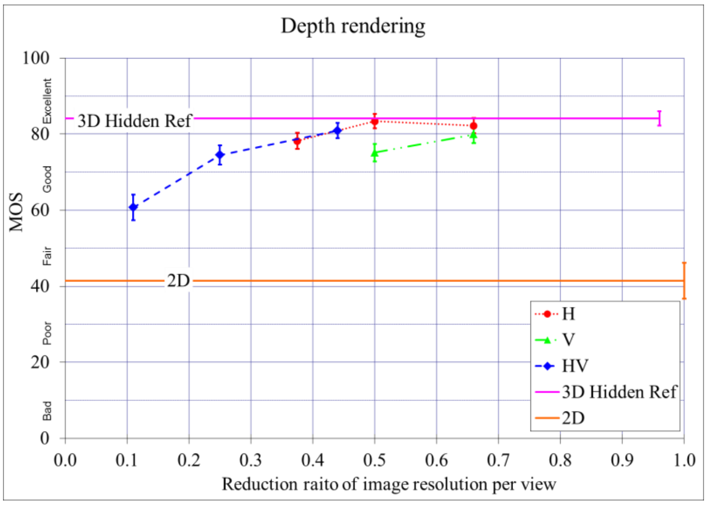


Figure 8-6 : MOS of depth rendering averaging over all the scenes (with their 95% confidence intervals) vs. Reduction ratio of Image resolution per view

Interlaced video vs. progressive video

In Figure 8-7, the MOS of visual experience for progressive video and interlace video are plotted separately. We can see that for progressive content, the visual experience of different types of resolution reduction maintains near "excellent" when resolution reduction ratio is larger than 0.5. With the same numbers of pixels, vertical and horizontal reductions maintain similar level of visual experience. For interlaced content, in case of horizontal resolution reduction, the curve has similar shape as the progressive content. However, in case of vertical resolution reduction, the visual experience reduces seriously with increasing resolution reduction. For instance, half horizontal reduction can still maintain a MOS more than 80, i.e., "excellent". However, half vertical reduction is rated only around 50 MOS as "fair". This effect is possibly related to the fact that the original resolution of interlaced content per view in vertical direction is only half of horizontal direction. Further resolution reduction in vertical direction might affect more in visual experience comparing to horizontal resolution reduction. It indicates that for playing interlaced content in interleaved displays, horizontal resolution reduction (as Side-by-Side) is a more suitable solution for maintaining the visual experience compared with vertical resolution reduction (as Top-and-Bottom).

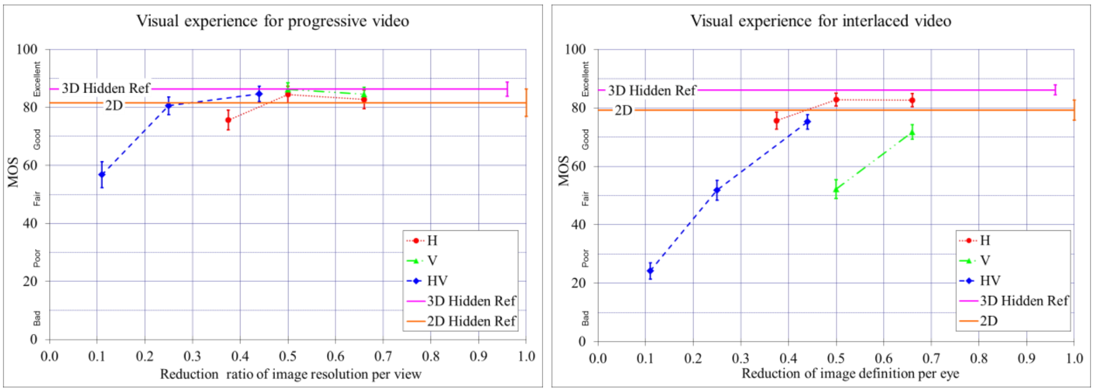


Figure 8-7 : MOS of depth rendering averaging over progressive contents (top) and interlaced content (bottom) (with their 95% confidence intervals) vs. Reduction ratio of image resolution per view

8.2.3 Discussion

The undistorted 2D content shows that the visual experience is lower than that of undistorted 3D content. It is easy for the user to distinguish the 2D and 3D content since 2D contain very poor depth rending.

The test results also demonstrate that reduction of definition in the left and right eye of a stereo image pair reduces the MOS of two QoE indicators: visual experience and depth rendering. It is therefore important to preserve the resolution of each image of a stereo pair to avoid the degradation of the user experience.

Moreover, the image format associated with the 3D display technique has a significant impact on user perception. For the interleaved display used in this study, in case of interlaced contents, horizontal resolution reduction provides better visual experience than vertical resolution reduction. Thus, it indicates that Side-by-Side (half) format is better than Top-and-Bottom format when playing interlaced content in interleaved displays. For progressive content, it seems that similar levels of visual experience between horizontal and vertical resolution reduction were rated in this study. However, since only low texture synthetic contents were used as progressive content, it might also relate to the texture level. Further studies with more stimuli in different levels of texture complexity are required to confirm this finding.

8.3 Experiment 2

The previous experiment showed that reducing resolution will reduce visual experience and depth rendering. Side-by-Side format is better than Top-and-Bottom format for interlaced content in interleaved displays since it can provide better visual experience. The rationale for Experiment 2 is to further understand the potential performance of frame compatible formats including Side-by-Side and Top-and-Bottom in the line interleaved display under different compression bitrates.

8.3.1 Methodology

This experiment was targeted to investigate the effect of compression bitrates (4 levels) and video formats (Side-by-Side, Top-and-Bottom and 2D HD) on visual experience of S-3DTV images. The method of this experiment is presented as follows:

1) **Stimuli:** four scenes were selected to cover different video complexities as follows:

- Foule (high, high, high): Many objects, some are highly textured, great depth range.
- Banc (high, high, high): Many objects, strong movement in the front plane, great depth range
- JT (low, low, low): Few objects and texture, little movement
- Tennis (mid, mid, mid): tennis player on the clay playground, middle level movement and mid-range of depth.

All the sources contents were interlaced content in 1080i25. Moreover, three types of video formats as 2D HD, Side-by-Side and Top-and-Bottom were considered in this experiment. The reference videos were provided in two full HD views (the left view and the right view) resolution. The 2D HD was the left view of undistorted 3D. Side-by-Side and Top-and-Bottom formats were converted from the reference 3D video. The conversion process was identical to the process present in Experiment 1 in this chapter. For each format, four compression bitrates including 5Mbps, 8Mbps, 12Mbps and 16Mbps were generated using a hardware based H.264 encoder and decoder (ATEME KFE SYSTEM, MPEG-4 AVC high profile and level 4.0). The video was first compressed and decompressed into raw video format to be played directly in the DVS as presented in experiment 1. Thus, for each scene, viewers were required to rate 14 stimuli as [3(video formats) x 4(levels of bitrates) + reference video + hidden reference video]. The hidden reference video was the same as the reference, the undistorted 3D video. Overall 56 stimuli were rated for each viewer in this experiment.

2) **Apparatus and test environment:** were identical to the Experiment 1 in this chapter.
3) **Observers:** were the same observers participated in the Experiment 1.
4) **Procedure**: were similar to the Experiment 1. SAMVIQ method was used as the main protocol. However, only one QoE indicator - visual experience was evaluated.

8.3.2 Result analysis

Figure 8-8 depicts the MOS of visual experience for all scenes for Side-by-Side, Top-and-Bottom and 2D HD formats in different compression bitrates. The level of visual experience may vary especially at a low bitrate among scenes depending on the scene complexity. However, the curves representing different video formats maintain the same shape and trend in different scenes. A one-way ANOVA (with Scene, compression bitrates, video format) was carried out on the raw subjective rating to test the main effects. The results revealed the main effect of bitrates ($p<0.02$) and video format ($p<0.001$). Scene ($p<0.60$) was not a significant factor.

Figure 8-9 depicts the MOS of visual experience averaged for all scenes. Similar to Experiment 1 in this chapter, even in high transmission bitrate such as 16Mbps, Top-and-Bottom format is rated only between “fair” and “poor” while Side-by-Side format and 2D HD format are both rated as between “excellent” and “good”. It confirms again the finding in Experiment 1 that Side-by-Side format performs better than Top-and-Bottom format for interlaced content in line interleaved display. The undistorted 3D video is rated as “excellent”, better than Side-by-Side and 2D HD in 16Mbps. The visual experience reduces with reduced compression bitrates for all the test formats. However, we can observe that the difference in visual experience between Side-by-Side and 2D HD increases with reducing compression bitrates. At low bitrates such as 5Mbps, 2D HD can still maintain around "good" quality while Side-by-Side is rated only as "poor".

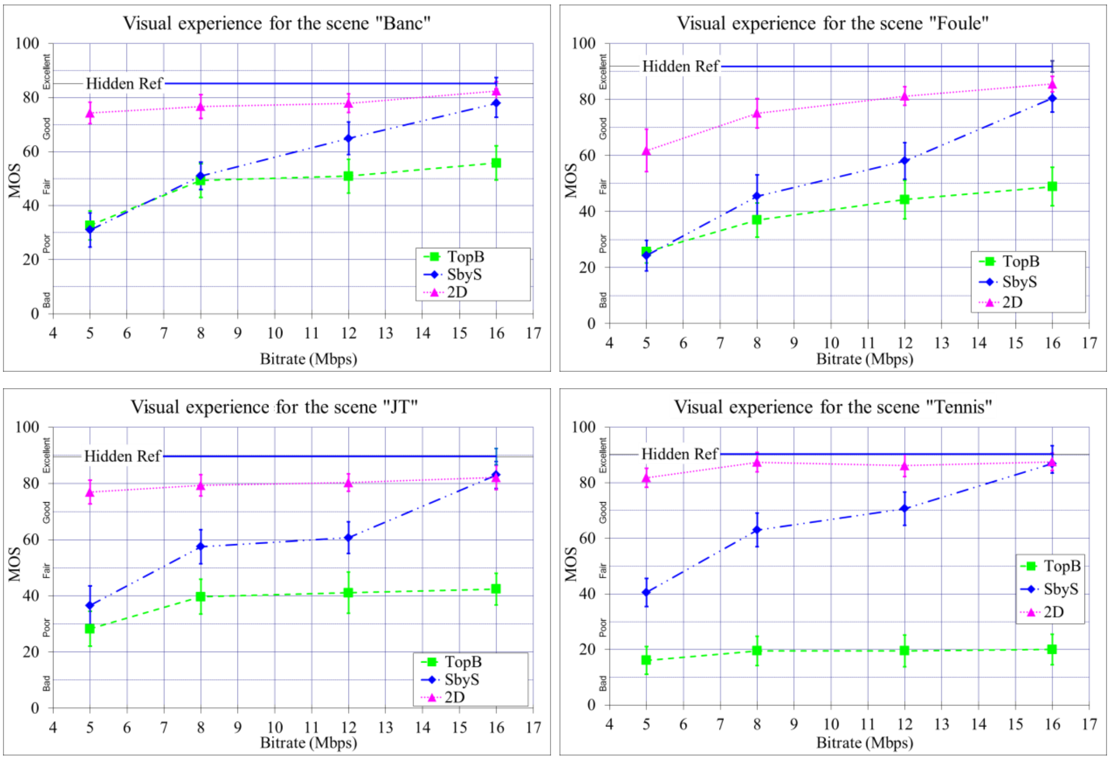


Figure 8-8 : MOS of visual experience for all scenes (with their 95% confidence intervals) vs. compression bitrates

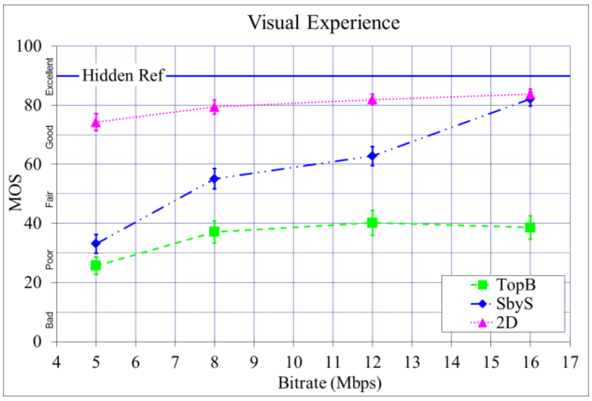


Figure 8-9 : MOS of visual experience averaging over all scenes (with their 95% confidence intervals) vs. compression bitrates

8.3.3 Discussion

Side-by-Side (half) format seems to be again more suitable for interlaced content in line interleaved display than the Top-and-Bottom one. It indicates that for achieving optimal visual experience, selecting the video format which fits to the 3D representation technique is very important. Moreover, in order to maintain the same visual experience as 2D, 3D may require higher bitrates. In this experiment, we can observe that at least 16Mbps is required for Side-by-Side format to reach similar level in visual experience as 2D.

8.4 Conclusion and recommendation

In this chapter, we designed two experiments to investigate the influence of different 3D representation formats on the QoE using line interleaved S-3DTV display. The first experiment concentrated in different formats of reduced definition. Second experiment was focused on different compression bitrates.

The results from the first experiment indicate that Side-by-Side format can provide better visual experience than Top-and-Bottom format for line interleaved display especially in case of interlaced scan content. It demonstrates that in order to optimize the quality of experiment of 3DTV, selection of 3D representation format should consider the interaction with 3D display technique.

The results from the second experiment show that in order to maintain the same level of visual experience, broadcasting 3D content in frame compatible format required more bitrate than 2D content.

Chapter 9 The impact of view asymmetry on the QoE of S-3DTV

Contents

9.1 Introduction

The view asymmetry problem in 3DTV can cause serious visual discomfort and therefore affect negatively the quality of experience. In (Kooi and Toet, 2004), the authors investigated the relative contribution of spatial imperfections in binocular image pairs that can cause viewing discomfort. Based on the subjective experiment, they estimated and proposed the threshold for different type of the binocular manipulation, e.g., rotation, magnification, vertical shift, luminance. Three main conclusions from their study are: 1) The factors that determine stereoscopic viewing comfort most strongly are vertical disparity, crosstalk and blur; 2) Individual difference of stereopsis only has very limited influence on the binocular viewing comfort; 3) Hyperstereopsis only has very weak effect on the visual discomfort. However, in their research, there might exist some potential problems: 1) The origin pair of image was acquired from cameras directly without any post-production, i.e. they might originally contain certain level of view asymmetries; 2) The presentation time was only 3 seconds per stimuli, i.e. it may be too short for viewer to review the whole image carefully; 3) The viewing environment, e.g., the viewing distance, the background light, the two projectors, were not following the standardized method.

In this study, we target to investigate the impact of view asymmetry on the QoE of 3DTV in a more critical way: (1) The origin image pairs are either synthetic content without any view asymmetries problem or natural content in which view asymmetries problem have been fixed by post production. The potential view asymmetries from the origin image pairs can be excluded in this study; (2) The subjective experiment strictly follows the standardized method in order to guarantee reliability and reproducibility of the experiment; (3) Both visual annoyance scale and visual comfort scale are used. The visibility and the visual annoyance thresholds as well as the visual comfort threshold are estimated.

This chapter is organized as follows. Firstly, three groups of view asymmetries consisting of luminance asymmetry (black and white), color asymmetry (red, green, blue), and geometrical asymmetry (vertical shift, rotation, magnification) are defined. Secondly, three pairs of stereoscopic images representing different texture level and binocular depth range (High texture and high depth, mid texture and mid depth, low texture and low depth) are selected. For each type of view asymmetries, four level of distortion are defined by expert test. Therefore, a large number of stimuli are generated. Finally, subjective quality assessments based on SAMVIQ method are conducted to evaluate the viewer's opinion in visual artifact (5 levels impairment scale) and visual comfort (5 levels continuous quality scale). The results show that every type of view asymmetries seriously induces visual artifacts and causes visual comfort if presented in a large enough amount. From the experimental data, the visibility threshold and the visual annoyance threshold as well as the visual comfort threshold are estimated. The thresholds obtained from this study allow a more accurate prediction of QoE from the specification of 3DTV system.

Being able to predict the level of visibility and annoyance of visual artifacts as well as visual discomfort from the specification of 3DTV helps the design and selection process. This study also provides the basis.

9.2 View asymmetry on 3DTV

The view asymmetry problem can be induced by difference sources. For example, from the content creation procedure, the toed-in camera configuration can produced vertical disparity and keystone distortion due to the geometry structure. The misalignment of camera position can result in vertical shift, rotation, magnification between views. The differences between camera focal lenses can provide different level of blur and magnification between views, the desynchronization of color or luminance on different camera sensors can induce color and luminance asymmetry. In the compression and transmission, asymmetry coding strategy can produce more visual artifacts on one view. Furthermore, in the final visualization part, the imperfection of filter in the display or the glasses can cause luminance asymmetry, color asymmetry or crosstalk. The misalignment of projectors position can also produce geometrical asymmetry.

In this study, we categorize the commonly encountered view asymmetries into three groups including the luminance asymmetry, color asymmetry and geometrical asymmetry in order to facilitate the analysis.

9.2.1 Luminance asymmetry

Luminance asymmetry is the most common asymmetry in 3DTV. It can be induced by the misalignment of the luminance level or the gamma function on cameras, the additional optics such as mirror rig on the camera system, the imperfection of the filter on the final display and glasses. To be more practical, the video engineers accustom to adapt the white level and black level in order to avoid the luminance asymmetry between cameras. Thus, luminance asymmetry can consist of two types of view asymmetries: the white level asymmetries and the black level asymmetries.

The white level asymmetry can be depicted as shown in Figure 9-1 and described by Equation 9-1.

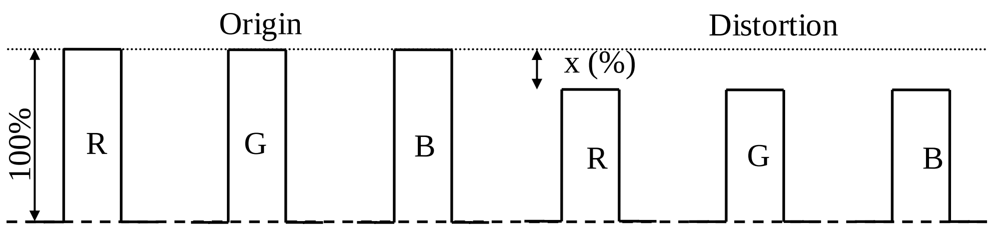


Figure 9-1 : white level asymmetry (distortion factor x percentage)

$$\boldsymbol{L_{distort}(RGB) = L_{origin}(RGB) \times (1 - x)} \quad \textbf{(9-1)}$$

with $L_{distort}, L_{origin}$ *and* x representing the origin luminance value of the image, the distorted luminance value of the image and the distortion level (percentage), respectively.

The black level asymmetry can be depicted as shown in Figure 9-2 and described by Equation 9-2.

$$\boldsymbol{L_{distort}(RGB) = L_{origin}(RGB) \times (1 - x) + 255 \times x} \qquad \textbf{(9-2)}$$

With $L_{distort}, L_{origin}$ *and* x representing the origin luminance value of the image, the distorted luminance value of the image and the distortion level (percentage), respectively.

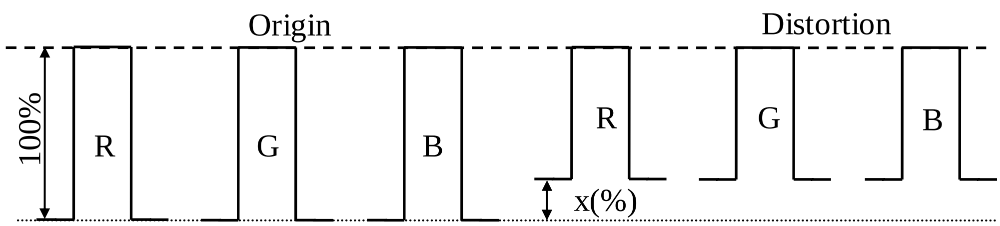


Figure 9-2 : Black level asymmetry (distortion factor x percentage)

9.2.2 Color asymmetry

The imperfection of filters on the stereoscopic production or viewing system can cause color asymmetry, e.g., the polarized filter in the display or the glasses. The imperfection adjustment of color triangle on the cameras may also induce color channel asymmetry. And of course, the color channels multiplex technique such as red/cyan anaglyph stereoscopic viewing technique cause serious color asymmetry problem. In this study, we stimulate the color asymmetry in a similar way as the white luminance asymmetry but limiting to one single color channel.

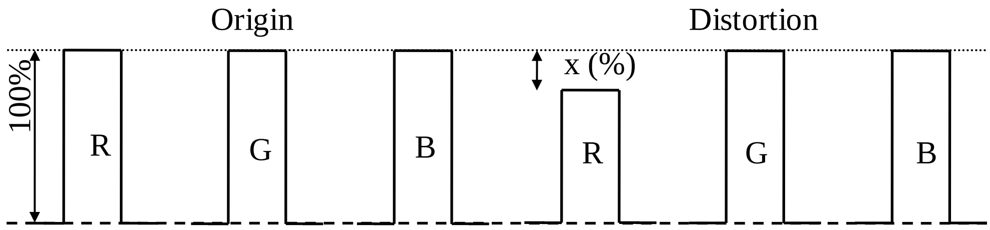


Figure 9-3 : Color asymmetry in Red channel (distortion factor x percentage)

Figure 9-3 illustrates the color asymmetry in Red channel. Green and Blue channel asymmetries are performed in a similar way and this procedure can be described by the below equation:

$$\boldsymbol{L_{distort}(R|G|B) = L_{origin}(R|G|B) \times (1 - x)} \qquad \textbf{(9-3)}$$

With $L_{distort}, L_{origin}\ and\ x$ representing the origin luminance value of the image in different color channel, the distorted luminance value of the image in different color channel and the distortion level (percentage), respectively. Four different levels of distortion for each color channel including 10%, 20%, 30% and 50% were selected as experimental stimuli for color asymmetry.

9.2.3 Geometrical asymmetry

The geometrical asymmetry can be induced by the camera configuration itself. In (Woods et al., 1993), the author analyzed the geometry of the stereoscopic camera and display systems in order to understand the effect of image distortion in stereoscopic video system. Their analysis pointed out that toed-in (converged) camera configuration can cause keystone distortion resulting in vertical disparity in the border of the view.

The geometrical asymmetries addressed in this study are more related to geometry misalignment of stereoscopic views due to positioning of the camera or projector or the inappropriate post-production. Three types of geometrical asymmetries are selected to be simulated including the vertical shift, rotation of one view and the magnification of one view. Figure 9-4 illustrates these three types of geometrical asymmetries.

The vertical shift of image only causes a uniform distribution of the vertical disparity. The rotation of one view and the magnification of one view induce both vertical disparity and unintentional horizontal disparity.

The rotation asymmetry can be denoted by the below equation:

$$\begin{bmatrix} \boldsymbol{d_{horizontal}} \\ \boldsymbol{d_{vertical}} \end{bmatrix} = \left\| \begin{bmatrix} \boldsymbol{cos\theta - 1} & \boldsymbol{-sin\theta} \\ \boldsymbol{sin\theta} & \boldsymbol{cos\theta - 1} \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix} \right\| \qquad \textbf{(9-4)}$$

with $d_{horizontal}, d_{vertical}, x\ and\ y$ representing the vertical disparity and horizontal disparity induced by rotation, the degree of rotation, and the position of the original pixel, respectively. This equation indicates that the distribution of vertical disparity and horizontal disparity for rotation asymmetry is not a linear function. The distortion level of both types of disparity increases from the rotation center to the border. The maximum horizontal and vertical disparity is located in the image border.

For the magnification asymmetry, the induced vertical and horizontal disparity can be expressed in the below equation:

$$\begin{bmatrix} \boldsymbol{d_{horizontal}} \\ \boldsymbol{d_{vertical}} \end{bmatrix} = [\boldsymbol{\mu - 1} \quad \boldsymbol{\mu - 1}] \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix} \qquad \textbf{(9-5)}$$

with $\mu, x\ and\ y$ representing the magnification ratio and the position of the original pixel. Both unintentional disparities increase from the image center to the image border. Thus, the maximum disparities are also located in the image border.

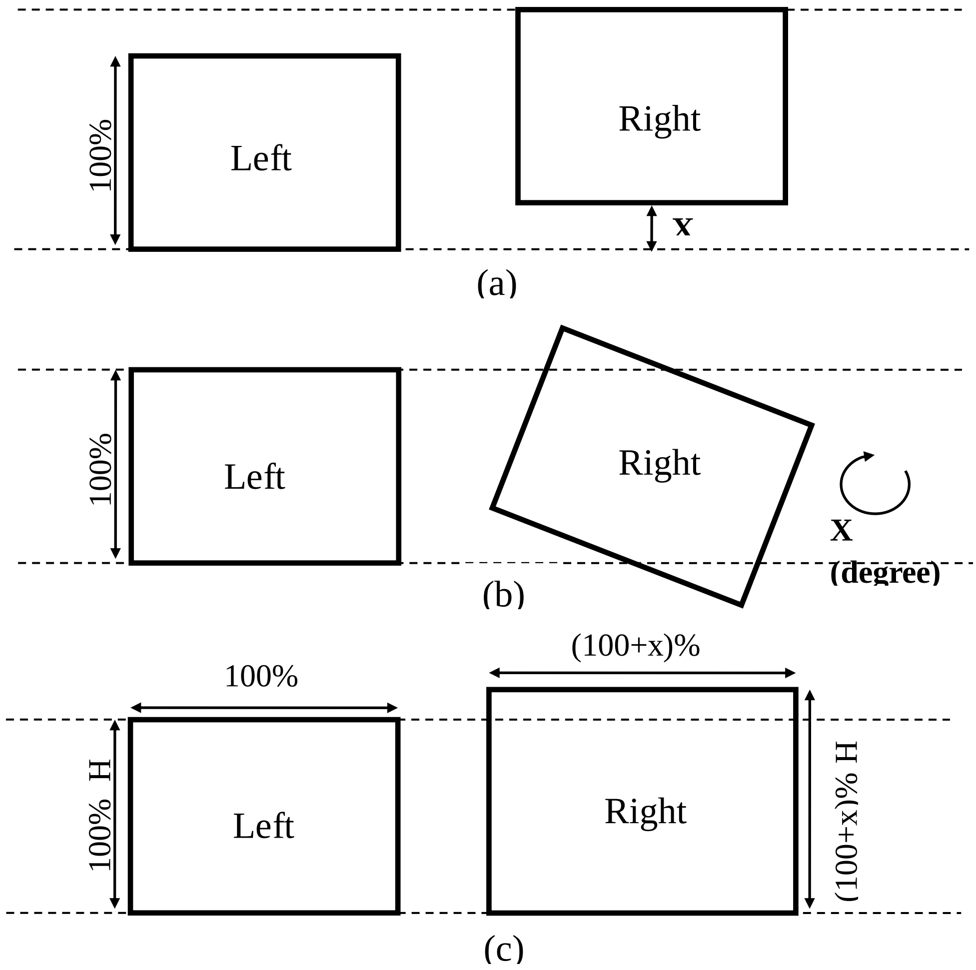


Figure 9-4 : Geometrical asymmetry (a) Vertical shift(x in percentage) (b) Rotation of one view (x in degree) (c) magnification of one view (x in percentage)

The simulation of geometrical asymmetries is implemented by Matlab. For simulating vertical shift, in the practical manipulation, in order to avoid the black border after vertically shift of the origin image, firstly a 110% resize function with lanczos3 filter were implemented on both views. After that, a simple crop function was used to crop the distorted image from the center of the resized image. Left and right views were both shifted in $\frac{x}{2}$ percentage of the height of the resized image in order to generate finally a x percentage distortion level vertical shift.

$$\begin{aligned} \boldsymbol{I_{left}^{distort}} &= \boldsymbol{Crop_{vertical}}(\boldsymbol{resize}\left(\boldsymbol{I_{left}^{origin}}, \frac{x}{2}\right)) \\ \boldsymbol{I_{right}^{distort}} &= \boldsymbol{Crop_{vertical}}(\boldsymbol{resize}\left(\boldsymbol{I_{right}^{origin}}\right), -\frac{x}{2}) \end{aligned} \qquad \textbf{(9-6)}$$

with $I^{distort}, I^{origin}$ *and* x denoting the distorted image, the origin image and the distortion level (in percentage of the width of the resized image).

For rotation asymmetry, it was similar to the manipulation of vertical disparity, resize function was implemented before the rotation in order to avoid the black border. It can be denoted as the below function:

$$\boldsymbol{I^{distort}} = \boldsymbol{Rotate}(\boldsymbol{resize}(\boldsymbol{I^{origin}}), \boldsymbol{x}) \qquad \textbf{(9-7)}$$

with $I^{distort}, I^{origin}$ *and* x denoting the distorted image, the origin image and the distortion level (in degree). The rotation function was integrated with a 'Bicubic' interpolation.

For magnification of one view, the procedure can be described as:

$$\boldsymbol{I^{distort} = Magnification(I^{origin}, 100 + x)} \qquad \textbf{(9-8)}$$

with x representing the distortion level in percentage of the height or width of the origin image. The magnification function was practically a resize function with a 'lanzcos3' filter.

9.3 Subjective QoE assessment

In this study, a subjective QoE assessment experiment is designed to assess the impact of view asymmetry on the QoE of 3DTV. Section 9.3.1 presents the experiment design. Section 9.3.2 focuses on result analysis.

9.3.1 General experiment design

1) **Stimuli:** three pairs of stereoscopic images representing different levels of image complexity were selected as the original images. Figure 9-5 depicts these three scenes. The Forest scene is a high-level texture scene with 0.2 diopters perceived depth range, the Butterfly scene is a middle-level texture scene with 0.1 diopters perceived depth range and the Basketball scene is low-level texture with 0 diopter (2D) perceived depth range which the right view is only the duplication of the left view. The synthetic scenes were rendered without any view asymmetry by Blender Software with a virtual parallel camera. The natural scene was acquired by a mirror rig toed-in camera. Post-processing was conducted to fix the possible view asymmetry by professional company. All the stereoscopic images were reviewed carefully by Pure software from Stereolab to guarantee that there were no potential view asymmetries left in the original image pairs.

Forest

Butterfly Basketball

Figure 9-5 : Three original scenes: Forest (high texture, 0.2 diopters depth), Butterfly (mid texture, 0.1 diopters depth) and Basketball (low texture, 0 diopter depth)

All the original images were manipulated to simulate the practical view asymmetry following the rules presented in Section 9.2. Thus, for each scene, eight types of view asymmetries were generated as shown in Table 9-1.

The selected distortion levels were decided by a pre-expert test. A large amount of distorted samples were generated in the pre-expert test, and three video experts decided the final four distortion levels in order to appropriately cover the assessment scales. Therefore, three scenes × eight types of view asymmetries × four levels of distortion, overall 96 view asymmetries samples were generated.

Table 9-1 : Eight types of view asymmetries with four-level distortion

Distortion level		Level 1	Level 2	Level 3	Leave 4
Luminance asymmetry	White	10%	20%	30%	50%
	Black	1%	5%	15%	25%
Color asymmetry	Red	10%	20%	30%	50%
	Green	10%	20%	30%	50%
	Blue	10%	20%	30%	50%
Geometrical asymmetry	Vertical shift	0.4%	1%	1.4%	1.8%
	Rotation	0.2 degree	0.5 degree	1 degree	2 degree
	Magnification	0.4%	1%	1.4%	2%

2) **Equipment:** the subjective assessment was conducted in a test room, which is compliant with the recommendations for subjective evaluation of visual data issued by ITU-R BT.500. A 46 inch line-interleaved stereoscopic television with 1920x1080 pixels was used as the final visualization terminal. The luminance, brightness, contrast and color of the display were adjusted to fulfill the normal gamma function (gamma equals 2.2), the PLUGE test from ITU-R BT.500 as well as the conventional color triangle. The crosstalk level was less than 3% and the luminance measured through glasses was more than 100cd/m^2. An additional 17 inch LCD display was used to display the interface for collecting the observer response. The viewing distance was fixed to 2.6 meter as 4.5 times of display height. A Quad-core 3000MHZ computer with HDMI 1.4 output and SEOVQ software was used to output the stimuli to the display and collect the observer responses.

3) **Observers:** 30 observers were recruited to participate in this test. All of them were non experts in the audiovisual and video domain. A vision test was performed on all testers to determine their visual performance and the potential impact on results. The test includes monocular visual acuity test, hyperopic trend, astigmatic trend, binocular distant vision acuity, dysphasia, fusion, stereoacuity and color vision. All observers had a normal or corrected to normal visual acuity and normal stereoacuity.

4) **Procedure:** written instructions detailing the task what the observers had to perform and the attributes they were asked to rate were given to the subjects before the start of the test. These instructions were then reiterated by the experimenter as to ensure the observer understand the task. SAMVIQ method was used to evaluate both the visual annoyance and visual comfort test. The experiment was separated into two parts. The first part consisting of eight tests (corresponding to eight types of view asymmetries) was targeted to assess the stimuli using a five-level impairment scale in terms of visual annoyance. The second part also consisted of eight tests (corresponding to eight types of view

asymmetries) but used a five-level continuous quality scale to access visual comfort. The five-level impairment scale and the five-level quality scales are depicted as the Table 9-2.

These two parts of the experiment were conducted in different days in order to avoid the accumulated visual discomfort as well as context effect. Moreover, all subjects were suggested to have a 5 minutes rest after every two tests and had a 10 minutes rest after every four tests.

For the visual annoyance test, there were 96 stimuli (three scenes × eight types of view asymmetries × four levels of distortion) presented to the subject. In order to facilitate the voting process in the impairment test of visual annoyance, each stimulus was presented for 18 seconds as a reference image pair following by a distorted image pair as shown in Figure 9-6.

Table 9-2 : Impairment scale and quality scale

Impairment scale (categorical)		Quality scale (continuous)	
5	Imperceptible	80-100	Excellent
4	Perceptible but not annoying	60-80	Good
3	Slightly annoying	40-60	Fair
2	Annoying	20-40	Poor
1	Very annoying	0-20	Bad

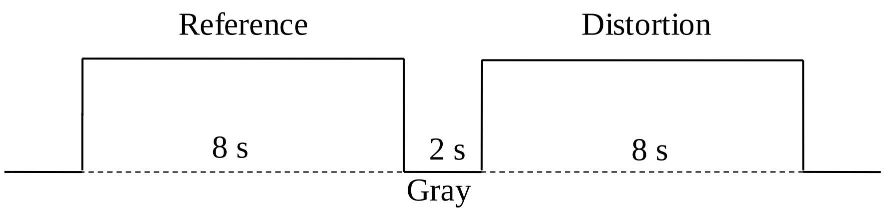


Figure 9-6 : Stimulus timeline for the visual annoyance test

For the visual comfort test, 144 stimuli were required (three scenes × eight types of view asymmetries × [four levels of distortion + empirical reference + hidden reference]) to vote. Each stimulus was an still stereoscopic image pair repeating for 8 seconds.

9.3.2 Result analysis

This study determines the relative importance of different types of view asymmetry to the quality of experience of 3DTV, mainly focusing on visual annoyance and visual discomfort.

First, following the recommendation from ITU-R BT.500, an approximation of the relationship between the MOS and the distortion level for each type of view asymmetries was implemented in order to find the relationship between the mean opinion score and the objective measure of the view asymmetries distortion levels.

Because all the distortion units were represented as a related unit, the symmetry logistic function was used to estimate the continuous relationship between the MOS vote and the distortion level as shown in the below equations:

$$\boldsymbol{V = 1 + \frac{4}{1+e^{\frac{a+D}{b}}}} \qquad \textbf{(9-9)}$$

$$\boldsymbol{V = \frac{100}{1+e^{\frac{a+D}{b}}}} \qquad \textbf{(9-10)}$$

The Equation 9-9 and 9-10 estimate the visual annoyance score and the visual discomfort score respectively with D representing the objective distortion level, V representing the MOS score, and a, b denoting the estimation constants. We used Matlab Curve fitting tools to estimate the parameters of the approximation, i.e., $a\ and\ b$. The R-square of the approximation for each asymmetry is more than 0.98 indicating the fit of the curve is robust and reliable.

Second, a visibility threshold was estimated following the ITU-BT 500 recommendation as the grade 4.5 between ‘imperceptible’ and ‘perceptible but not annoying’ in the minimum curve (the lower bound of the MOS with confidence interval) of the impairment scale. In addition, a visual annoyance threshold was defined similarly as the grade 3.5 between ‘perceptible but not annoying’ and ‘slightly annoying’ in the minimum curve. For the visual comfort scale, we used the grade 60 (representing 80 percent of people accept this level of visual comfort) as a threshold for the visual comfort following our previous recommendation as shown in Figure 9-7. The relationship between the different thresholds especially between the visual annoyance and visual comfort will be discussed later.

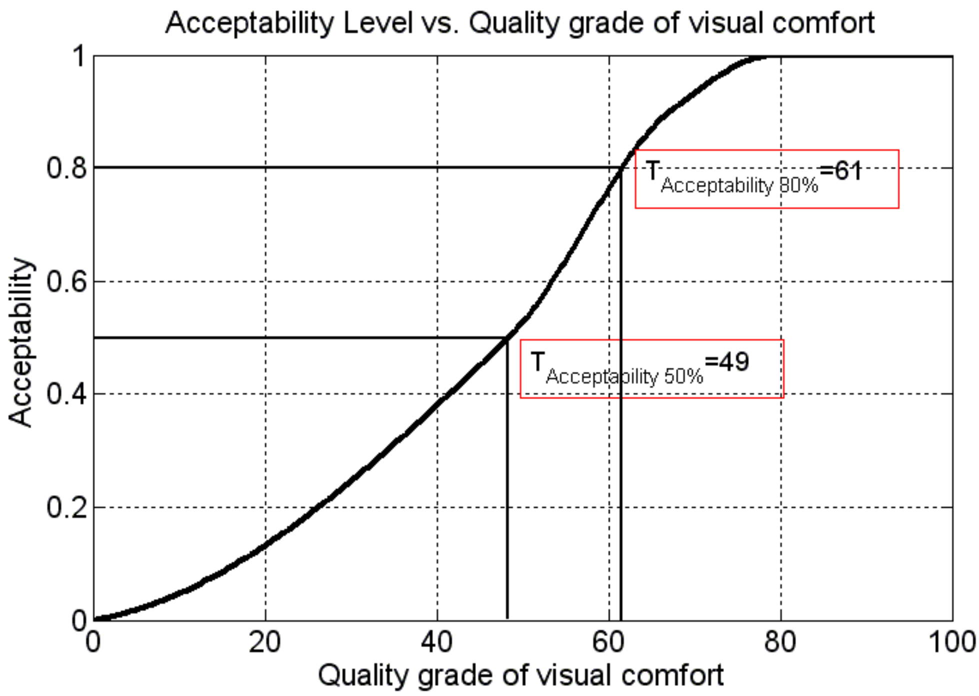


Figure 9-7 : Acceptability versus quality of visual comfort (from Figure 6-4, Chapter 6)

9.3.2.1 Luminance asymmetry

Figure 9-8 depicts the MOS scores of visual annoyance and visual comfort versus distortion level of black and white level asymmetry with visibility threshold, visual annoyance threshold for visual artifact as well as acceptability threshold for visual comfort. The subjects can perceive minimum 3 percent of black level asymmetry and start to feel annoyance until reaching 15 percent. The acceptability threshold for visual comfort is 11 percent between the visibility threshold and visual annoyance threshold. The subjects were less sensitive for the white level asymmetry compared with the black level asymmetry resulting in 11 percent for visibility threshold, 27 percent for visual annoyance threshold as well as 20 percent for acceptability threshold of visual comfort.

The one way ANOVA analysis shows that the variation of both white level and black level are significant for visual annoyance and visual comfort ($p<0.001$). Increasing the distortion level of course increases the visual annoyance level and reduces the visual comfort. However, image complexity is insignificant for the change of MOS scores in both white and black level asymmetry test ($p<0.97$).

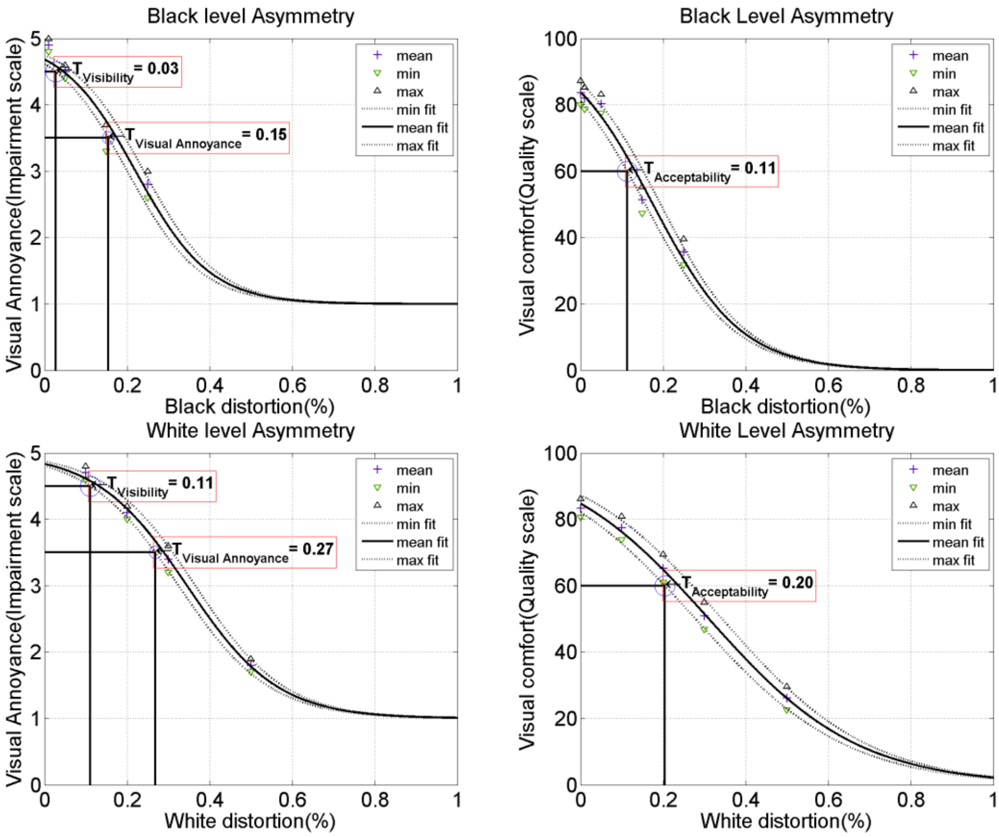


Figure 9-8 : The MOS score of visual annoyance (left column) and visual comfort (right column) with 95% confidence interval vs. Distortion level of Black (top) and White level (bottom) asymmetry: visibility threshold ($T_{Visibility}$), visual annoyance threshold ($T_{Visual\ Annoyance}$) and acceptability threshold ($T_{Acceptability}$)

9.3.2.2 Color asymmetry

It is important to clarify that the subjects with color deficiency were rejected since the color deficiency may affect their perception for different color channel asymmetry.

The MOS score with 95 percent confidence interval versus color asymmetry distortion is plotted in Figure 9-9.

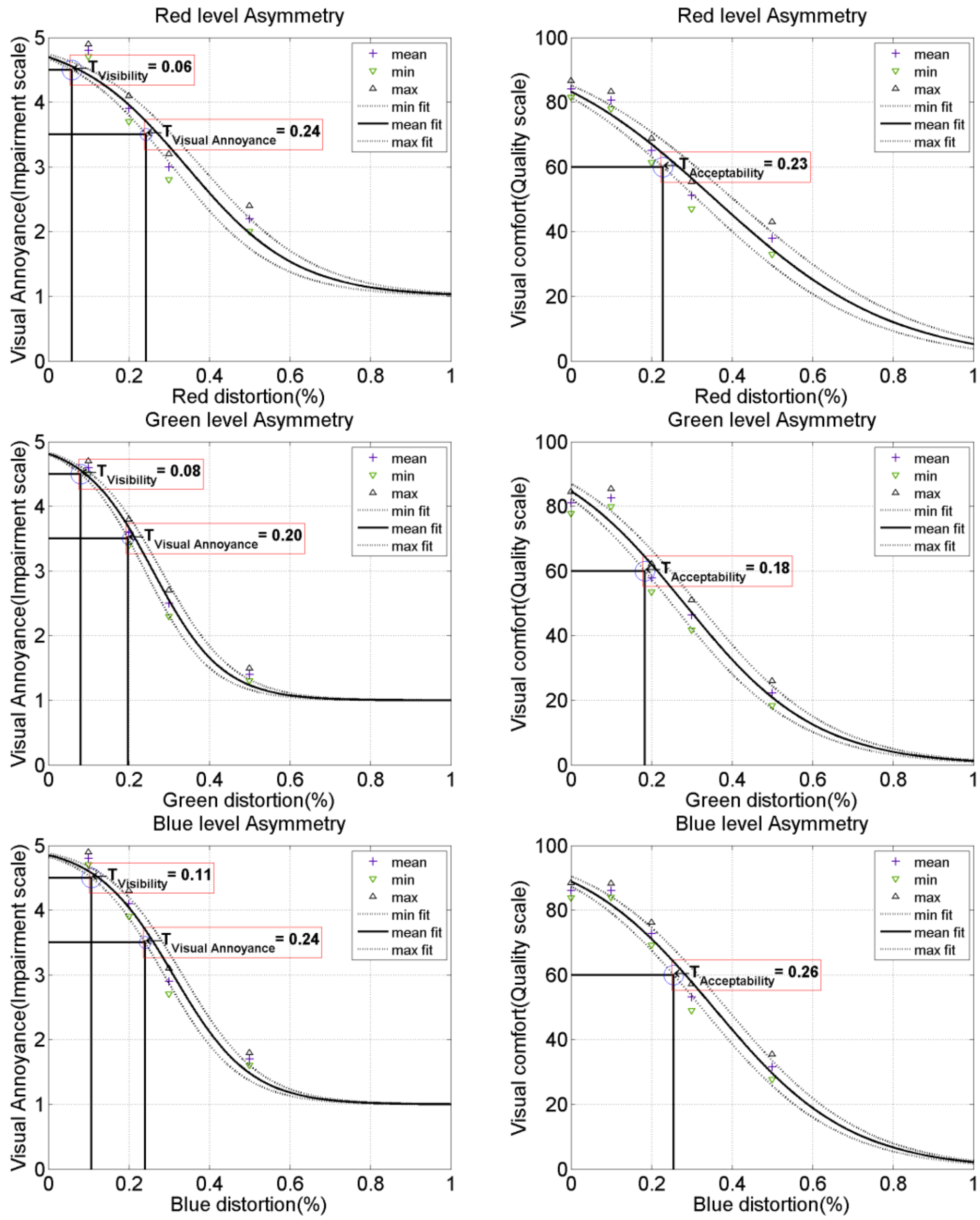


Figure 9-9 : The MOS score of visual annoyance (left column) and visual comfort (right column) with 95% confidence interval vs. Distortion level of Red (top), Green (mid), Blue (bottom) level asymmetry: visibility threshold ($T_{Visibility}$), visual annoyance threshold ($T_{Visual\ Annoyance}$) and acceptability threshold ($T_{Acceptability}$)

The visibility thresholds shown in Figure 9-9 are 6 percent, 8 percent and 11 percent for red, green, blue color respectively. Considering the approximation error (for red and green asymmetry, the curve is underestimate for small distortion level which can be observed from Figure 9-9 top left and mid left), 10 percent can serve as a more general visibility threshold for all color channels. Similarly, from the observation of the approximation of curve which likely indicates 24 percent, 20 percent and 24 percent for the visual annoyance threshold of red, green and blue level asymmetry respectively, we suggest that 20 percent can be used as a common threshold. The acceptability thresholds for visual comfort are located near the visual annoyance threshold and likely serve as a common threshold as 20 percent. Statistical analysis shows that the MOS score reduces significantly with the increasing of color asymmetry. However, in the color asymmetry test, image complexity is not a significant factor of visual annoyance and visual comfort (p<0.97). Moreover, different color components themselves, i.e., red, green and blue are insignificant for the MOS score for visual comfort and visual annoyance (p<0.87).

9.3.2.3 Geometrical asymmetry

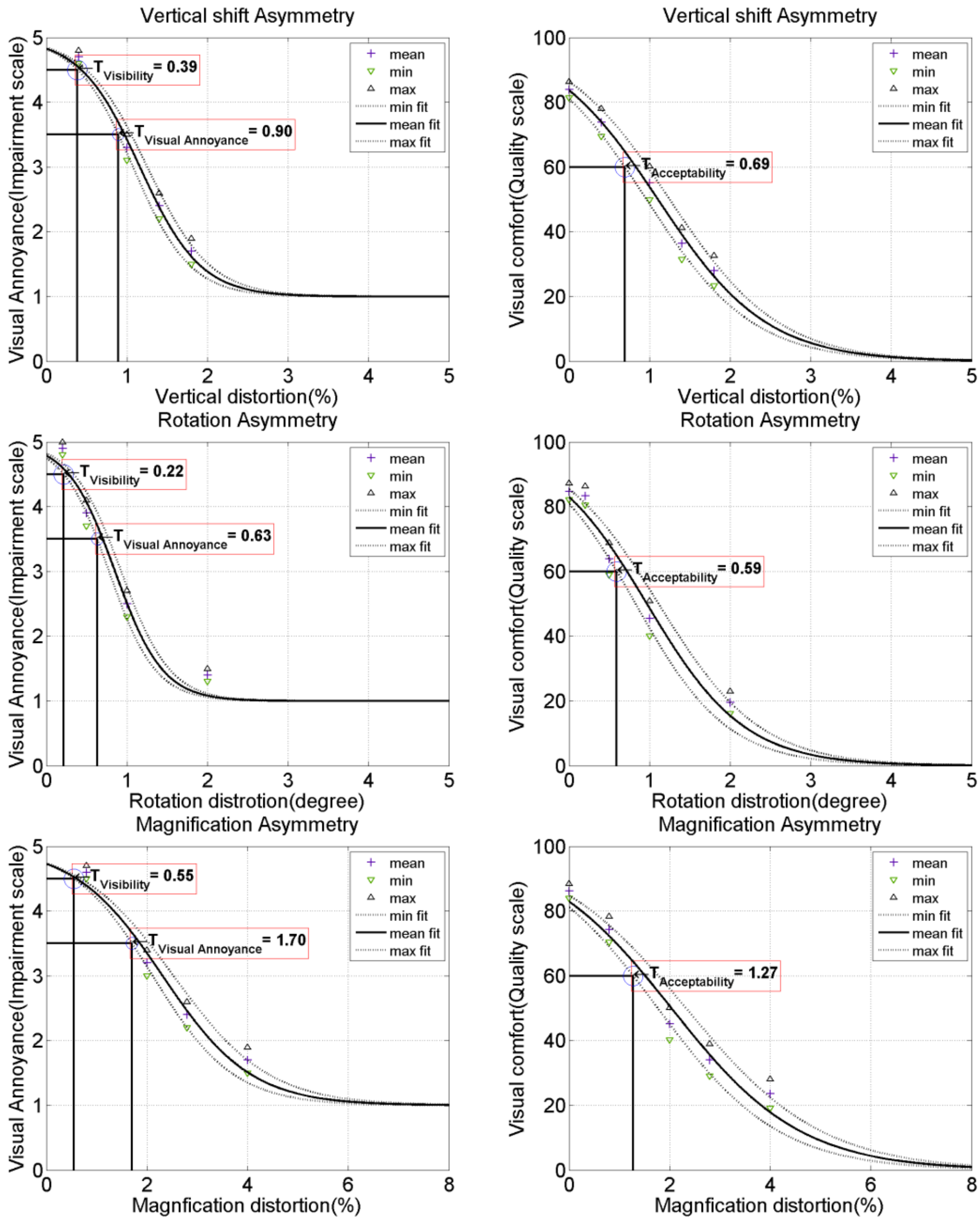


Figure 9-10 : The MOS score of visual annoyance (left column) and visual comfort (right column) with 95% confidence interval vs. Distortion level of Vertical shift (top), Rotation (mid), Magnification (bottom) asymmetry: visibility threshold ($T_{Visibility}$), visual annoyance threshold ($T_{Visual\ Annoyance}$) and acceptability threshold ($T_{Acceptability}$)

The results of geometrical distortion are depicted in Figure 9-10. The subjects can perceive 0.39 percent of vertical shift asymmetry, 0.22 degree of rotation asymmetry and 0.55 percent magnification asymmetry. Visual annoyance will be likely reported in 0.9 percent, 0.63 degree and 1.7 percent for vertical shift, rotation and magnification asymmetry respectively. People were likely to accept the visual comfort level minimally in 0.69 percent for vertical shift, 0.59 percent for rotation and 1.27 percent for magnification which are located between the visibility threshold and visual annoyance threshold.

In Section 9.2.3, we had already discussed the distribution of unintentional vertical disparity and horizontal disparity caused by different types of geometrical distortions. Vertical shift induces a uniform level of vertical disparity in the whole image while rotation and magnification bring both unintentional vertical and horizontal disparities. The amount of vertical and horizontal disparities increases from 0 in the image center to maximum in the image border. The maximum vertical disparity of rotation asymmetries locates in the four corners of the image (top left, top right, bottom left, bottom right) while for magnification asymmetry, maximum vertical disparity can be perceived in the top border and bottom border in the image. The maximum unintentional horizontal disparity can be found in the vertical center axis in the case of rotation and in the left/right border in the case of magnification. The unintentional horizontal disparity will pollute the original binocular disparities information resulting in visible artifact in depth. Table 9-3 present the maximum and average unintentional horizontal and vertical disparities in unit of pixel for each threshold of each type of geometrical asymmetry.

Table 9-3 : Unintentional vertical and horizontal disparities for each threshold and each type of geometrical asymmetry (pixel unit)

		Vertical shift		Rotation		Magnification	
		V	H	V	H	V	H
$T_{Visibility}$	Max	4.2	0	3.7	2	2.9	5.3
	Mean			1.8	1	1.5	2.6
$T_{Visual\ Annoyance}$	Max	9.7		10.6	5.6	9.2	16.3
	Mean			5.3	2.7	4.6	8.2
$T_{Acceptabilit\ y}$	Max	7.5		9.8	5.5	6.9	12.2
	mean			4.9	2.7	3.4	6

The maximum vertical disparities for visibility threshold of vertical shift, rotation and magnification are 4.2 pixels, 3.7 pixels, and 2.9 pixels, respectively. The differences of maximum vertical disparities among three types of view asymmetries can be likely explained by that the unintentional horizontal disparities. 2 pixels for rotation asymmetry and 5.3 pixels for magnification asymmetry induced more visual artifacts while for vertical shift case there is no unintentional horizontal disparity. Interesting, for visual annoyance threshold, the maximum vertical disparity for these three types of geometrical asymmetry serves in a similar amount around 10 pixels although the maximum horizontal disparity perform quite differently (0 pixel for vertical shift case, 5.6 pixels for rotation and 16.3 pixels for magnification). It may indicate that the visual annoyance mainly depends on the maximum vertical disparity. In case of visual comfort threshold, the maximum vertical disparity is around 7 to 9 pixels.

ANOVA analysis confirmed that all types of geometrical asymmetries significantly increase the visual annoyance and reduce the visual comfort ($p<0.001$). The MOS results from different scene did not perform significantly difference at the same level of distortion ($p<0.95$) for all types of geometrical asymmetries.

9.4 Conclusion and recommendation

In this study, we designed a subjective quality experiment based on the SAMVIQ method to measure the impact of view asymmetry on visual annoyance and visual comfort of stereoscopic still images. Our result shows that all types of view asymmetry increase visual annoyance and reduce visual comfort if a large amount is induced. However, it is possible to avoid the visual annoyance and visual discomfort problem if view asymmetries maintain within certain amount. Three thresholds including the visibility threshold, visual annoyance threshold as well as more practically, the acceptability (80 percent of viewers accept the visual comfort level) threshold were estimated in this study. Table 9-4 summaries estimated thresholds for view asymmetries. For three types of geometrical asymmetry, we found out that the maximum vertical disparity can be used as a common indicator since people are likely more sensitive to the maximum vertical disparity.

Table 9-4 : Estimated thresholds for view asymmetries

	Visibility threshold	Visual annoyance threshold	80% Acceptability threshold
Luminance asymmetry			
Black level	3 %	15%	11%
White level	11%	27%	20%
Color asymmetry			
R,G,B level	10%	20%	20%
Geometrical asymmetry			
Vertical disparity	0.39 %	0.9%	0.69%
Rotation	0.22 degree	0.63 degree	0.59 degree
Magnification	0.55 %	1.7%	1.27%
maximum vertical disparity	2.8 arcmin	7 arcmin	5.6 arcmin

Most of estimated thresholds presented in this study are more critical compared with the proposed thresholds in (Kooi and Toet, 2004). It can be explained by: first, the stimuli were carefully selected in order to cover the evaluation scale more evenly as well as more levels of distortion for each type of view asymmetries were measured in our test. Thus, more precise threshold can be proposed, e.g., 2.8 arcmin for vertical shift compared to roughly less than 34 arcmin (1 PD) proposed in (Kooi and Toet, 2004). Second, the presentation time for each stimulus was 8 seconds as well as SAMVIQ method allowed viewer to review the samples freely. However, in (Kooi and Toet, 2004), there was only 3 seconds of presentation time for each stimulus. Longer presentation and more freedom to review the image may allow viewer to be more critical of the visual artifact. Thus, our results are more accurate and reliable.

The collected results from this study and the proposed threshold allow a more precise prediction of visual annoyance and visual discomfort problem of a stereoscopic imaging system. It will certainly help the design and selection of stereoscopic imaging system. The visibility threshold can be used to guide the optimal stereoscopic imaging system design. More practically, the 80% acceptability threshold can be used for the recommendation of 3DTV services due to the reason that there was no optimal system in the market. Thus, we recommend the view asymmetry of stereoscopic imaging system:

- **The black level asymmetry should be less than 11% and the white level should be less than 20%**
- **For color asymmetry, less than 20% asymmetry should be required.**
- **For the maximum unintentional vertical disparity, it should be less than 5.6 arcmin.**

General Conclusion

I. Main conclusions

Chapter 1 investigated the QoE challenges for S-3DTV. The challenges can be divides into two levels: perceptual level and technical level. In perceptual level, S-3DTV might enhance the depth perception since the added binocular depth information in the viewing distance of S-3DTV applications is a sensitive depth cue. However, this enhancement of depth is only an illusion. The discrepancies (e.g., isolation of accommodation and convergence) between viewing S-3DTV and actually viewing the real scene might cause visual discomfort and visual fatigue problems. In technical levels, by reviewing different techniques in different individual parts of the S-3DTV broadcasting chain, the conclusion is that there are no transparent techniques. Many technical issues exist in different techniques and might have potential impact on the final QoE.

Chapter 2 reviewed state-of-art of subjective QoE assessment for S-3DTV focusing on ITU recommendations and explorative studies in the literature. We revealed that ITU recommendations such as ITU-R BT.500 are not sufficient to evaluate the 3D QoE as many new characteristic of S-3DTV are not considered. Concerning explorative studies, we revealed two main problems as the lack of clear definition of QoE indicators and the specification of the viewing environments (or “common feature”) for subjective QoE assessment for S-3DTV. Based on the above understanding, we proposed two main points for developing new subjective QoE assessment for S-3DTV: first, the QoE of S-3DTV is multidimensional. Six possible indicators for assessing short-term effect of QoE (2D image quality, depth quantity, visual comfort, depth rendering, naturalness and visual experience) were summarized and defined. One particular QoE indicator, visual fatigue, was defined for assessing long-term effect of QoE; second, new factors affecting QoE assessment of S-3DTV, e.g., specification of viewing environment, were addressed and discussed.

Chapter 3 focused on one of the new factors required to be specified for subjective QoE assessment of S-3DTV: characterizing S-3DTV display in term of luminance rendering performance and depth rendering performance. For luminance rendering, we proposed new characteristics of S-3DTV display to be measured and to be specified into the requirement of subjective QoE assessment. A case study comparing different luminance rendering performance among different S-3DTV displays was presented. The results showed that the luminance reduction ratio, the final perceived luminance, the gamma function and the crosstalk vary among different displays. These differences might have a potential impact on the final QoE and require future study to confirm. For this thesis, the strategy is to select the best display in terms of best luminance rendering performance taking into account the case study. For depth rendering, we defined depth rendering ability as the amount of rendering depth plane in a comfortable viewing zone for S-3DTV display, combining the physical parameter for depth geometry and perceptual parameters for visual comfort. We compared the depth rendering ability among different display techniques. The result analysis revealed that the depth rendering ability of S-3DTV display mainly depends on the viewing distance, the pixel size, the screen size and the organization of screen pixel for views.

Chapter 4 designed an experiment to measure visual fatigue for viewing one hour's video content in optimal viewing condition of S-3DTV. Both 2D and 3D contents were used in this experiment in order to understand whether there are differences concerning visual fatigue between 2D and 3D visualization. Three methods including questionnaire, vision test and EEG measurement were used to indicate visual fatigue. Concerning the questionnaire and the vision test, the result shows that there were no significant differences concerning the reported level of visual fatigue between 2D and 3D viewing. Thus, we concluded that viewing S-3DTV in optimal viewing condition might not result in higher level visual fatigue than viewing traditional 2DTV. However, the result of EEG measurement indicated significant difference especially in the power strength of beta and gamma band in most of the frontal and posterior of cerebrum. However, this might not necessarily related to visual fatigue. This might only reflect the difference in brain process related to depth perception between 2D and 3D viewing.

Chapter 5 proposed shooting rules to optimize the QoE of S-3DTV by considering stereoscopic distortion and constraint of comfortable viewing. Stereoscopic shape distortion was defined based on a geometry model mapping the binocular depth perception from the camera space to the visualization space. Several different camera models and configurations were analyzed. A comfortable viewing zone was summarized combining the proposed thresholds in the literatures. Our proposal of improved shooting rules consists of three points: 1) to adapt the camera parameters or scene parameters to avoid stereoscopic distortion; 2) to adapt camera parameters or scene parameters to guarantee that the perceived binocular depth is maintained in the comfortable viewing zone; 3) to guarantee that optimizing visual comfort is prior to optimizing stereoscopic distortion. A subjective QoE experiment was designed to judge the proposed shooting rules using three QoE indicators (visual experience, visual comfort and depth rendering). The results showed that the proposed shooting rules and associated priorities can ensure an optimized QoE.

Chapter 6 aimed to explore the impact of variation of perceived depth rendering on the QoE of S-3DTV. The results showed that 1) increasing the perceived binocular depth increases the depth quantity but decrease the visual comfort; 2) 2D image quality is not affected by variation of perceived binocular depth; 3) high level QoE indicators including depth rendering, naturalness and visual experience can be predicted by a weighted sum of image quality, depth quantity and visual comfort. Moreover, recommendations for the maximum disparity level of content production are proposed as 0.2 DOF for synthetic contents and 0.1 DOF for natural contents.

Chapter 7 investigated the effect of JPEG 2000 compression on the QoE of S-3DTV. The results showed that compression distortion induces a global degradation for all the QoE indicators. Moreover, comparing the effect of compression on 2D and 3D images, 3D shows advantages in visual experience in case of without any compression or in low compression ratios. However, these advantages will disappear with increasing compression distortion. The result might be due to the stereoscopic artifacts. The recommendation from this study is that there is no interest to provide 3D service in low bitrates as it provides even worse visual experience than 2D.

Chapter 8 explored the impact of image representation formats on the QoE of line interleaved display by two experiments. The results from the first experiment revealed that the cooperation between the 3D representation format and the S-3DTV display technique has impact on the QoE of S-3DTV. Side-by-Side format is better adapted with line-interleaved display than Top-and-Bottom format. The second experiment was designed to compare frame compatible formats (Side-by-Side, Top-and-Bottom) using different transmission bitrates. The results confirm the advantage of the Side-by-Side format to the Top-and-Bottom format in the same compression bitrate. Moreover, the results also indicated that to maintain the same visual experience as 2DHD, 3D required higher bitrates. The recommended bitrate for broadcasting the 3D Side-by-Side format using H.264 compression is at least 16Mb/s.

Chapter 9 focused on measurement of the impact of view asymmetry on the QoE of S-3DTV. Three groups (overall eight types) of view asymmetries including luminance, color and geometry were simulated. Our results confirmed that view asymmetries induce visual annoyance and cause visual discomfort. In order to optimize the QoE of S-3DTV, we recommended the view asymmetry of stereoscopic imaging system should be fulfilled: 1) The black level asymmetry should be less than 11% and the white level should be less than 20%; 2) For color asymmetry, less than 20% asymmetry should be required; 3) For the maximum unintentional vertical disparity, it should be maintain less than 5.6 arcmin.

II. Contributions

The contribution of this thesis covers three levels as the aim of this thesis:

- The first level is to develop new methodologies to assess 3D QoE. We propose to use multi-dimensional QoE indicators to measure the QoE of S-3DTV. We also highlight and reveal new factors affecting the 3D QoE in subjective QoE assessment. Part of these proposals have been submitted and accepted as contributions in ITU draft recommendation ITU-T P.3D-sam for subjective assessment methods for 3D video quality.
- The second level is to understand the impact of perceptual and technical issues on the QoE of S-3DTV. Several question concerning the content acquisition, 3D representation format, image compression and transmission bitrate have been addressed and been investigated.
- The third level is to provide recommendations to optimize the QoE of S-3DTV as follows:
  - Shooting rules to optimize content acquisition for S-3DTV. This work was granted as a France and International patent.

- Depth budget for synthetic contents production and natural contents production to guarantee visual comfort.
- Side-by-Side as the appropriate frame-compatible format for line interleaved display.
- Higher bitrate to broadcasting 3D content in frame-compatible format than 2D HD content.
- Visibility thresholds and visual comfort thresholds for luminance, color and geometrical asymmetries.

III. Perspective

Future research will focus on modeling the 3D QoE. We make a very first step as proposing that high level QoE indicator (visual experience, depth rendering and naturalness) can be estimated as a weighted sum of basic level QoE indicators (2D image quality, depth quantity and visual discomfort) in Chapter 6 and Chapter 7. However, only single factor variation is considered in each chapter (perceived depth variation in Chapter 6 and compression distortion in Chapter 7). In order to provide a more general QoE model and to reveal the relationship between the different QoE indicators, future research combining multi factors affecting the QoE of S-3DTV is required. Modeling of the 3D QoE will lead to a proposal of objective metric for measuring the S-3DTV QoE.

Résumé en Français

But de la thèse

La télévision 3D stéréoscopique (TV S-3D) pourrait être le successeur de la TV HD. Comparée à la TV 2D conventionnelle, l'intérêt de la TV S-3D est de fournir aux observateurs une sensation de profondeur accrue. Cependant, la TV S-3D ne permet pas une représentation parfaite de la réalité ; il s'agit seulement d'une illusion issue de deux images planes. Ainsi, les nouvelles questions propres à la TV S-3D comme l'inconfort visuel ou la déformation du relief pourraient être induites par des problèmes perceptuels et/ou techniques.

La qualité d'expérience (QoE) est une mesure subjective de l'expérience client d'un service. La « qualité d'image » est souvent employée pour représenter la QoE en TV 2D. L'évaluation subjective de la qualité est le moyen conventionnel utilisé pour évaluer la « qualité d'image » d'un système TV 2D. Cependant, la « qualité d'image » n'est pas suffisante pour représenter la QoE en TV S-3D parce qu'elle ne peut pas directement mettre en avant les avantages (par exemple une perception de la profondeur améliorée) et les problèmes (par exemple un inconfort visuel) liés à la TV S-3D. De plus, les méthodes conventionnelles d'évaluation subjective de la qualité ne prennent pas en compte les nouvelles caractéristiques de la TV S-3D, comme l'illustre le manque de spécifications de l'environnement de visualisation. Ainsi, le développement de nouvelles méthodes d'évaluation subjective de la QoE est indispensable pour caractériser la QoE de la TV S-3D, faciliter les spécifications des architectures de bout en bout et optimiser la conception des techniques de diffusion de la TV S-3D.

Le but de cette thèse est triple :

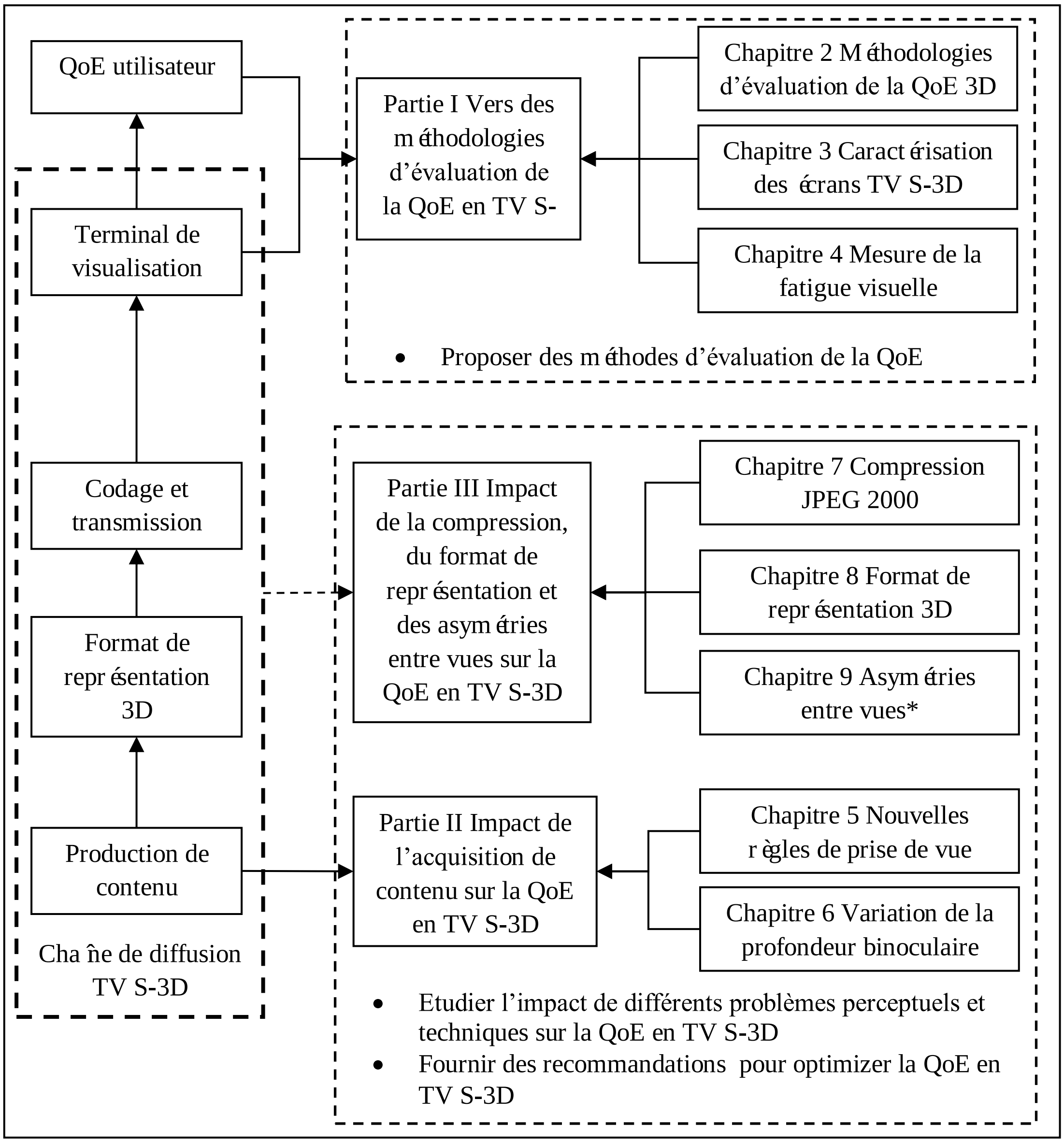
- Proposer une nouvelle méthode pour évaluer la QoE en TV S-3D
- Utiliser la méthode proposée pour étudier l'impact de différents problèmes perceptuels et techniques (le long de la chaîne de diffusion) sur la QoE en TV S-3D
- Fournir des recommandations perceptuelles et technologiques destinées à optimiser la QoE en TV S-3D

Vue d'ensemble de la thèse

Le chapitre 1 présente les défis liés à la QoE en TV S-3D comme base de cette thèse. En introduisant les fondements de la perception de la profondeur et les principes des systèmes vidéo stéréoscopiques, nous présentons les avantages fondamentaux (perception de la profondeur améliorée) et les problèmes (inconfort et fatigue visuels) liés à la QoE en TV S-3D. De plus, les questions liées à la QoE sont présentées et discutées en prenant en considération les différents éléments de la chaîne de diffusion TV S-3D (production de contenu, format de représentation 3D, codage et transmission et terminal de visualisation).

Les contributions de cette thèse sont organisées en trois parties indépendantes comme l'illustre la Figure R- 1. Chaque partie correspond à différents éléments de la chaîne de diffusion en TV S-3D.

La partie I, composée de trois chapitres (chapitres 2, 3 et 4), présente les contributions de la thèse sur l'évolution des méthodologies destinées à évaluer la QoE 3D. Tout d'abord, dans le chapitre 2, nous passons en revue les recommandations de l'UIT et les études exploratoires liées à l'évaluation subjective de la QoE en TV S-3D. Ensuite, pour adapter l'évaluation subjective de la QoE dans son ensemble, nous proposons d'employer des indicateurs multidimensionnels de la QoE et de considérer de nouveaux facteurs impactant la QoE. L'évaluation subjective de la QoE avec des indicateurs multidimensionnels servira de méthode principale pour l'évaluation de la QoE dans cette thèse. D'autre part, lors d'un test subjectif, la performance des écrans est un point clé, dont l'impact sur la QoE en TV S-3D est réel. C'est pourquoi nous proposons dans le chapitre 3 une méthode de caractérisation du rendu de la luminance et de la profondeur. Enfin, le chapitre 4 présente une étude de mesure de la fatigue visuelle dans des conditions de visualisation optimales. Trois méthodes, comprenant un test de vision, un questionnaire et des mesures EEG, sont employées pour mesurer la fatigue visuelle dans cette étude.



* L'asymétrie entre vues est un problème global lié à chaque partie de la chaîne TV S-3D.

Figure R- 1 : Vue d'ensemble des contributions de cette thèse

La partie II, composée des chapitres 5 et 6, présente les contributions de la thèse destinées à comprendre l'impact de l'acquisition de contenus sur la QoE en TV S-3D. Dans le chapitre 5, nous proposons des règles de prise de vue stéréoscopiques pour optimiser l'acquisition de contenus en TV S-3D, en prenant en considération les déformations stéréoscopiques et la zone de confort visuel dans la perception finale. Pour cela, des contenus synthétiques sont générés selon différentes conditions correspondant aux règles de prise de vue optimales proposées. Ensuite, une évaluation subjective est réalisée avec trois indicateurs de QoE, afin de vérifier les règles de prise de vue optimales proposées. Le chapitre 6 présente la mise en œuvre et les résultats d'une évaluation subjective, utilisant six indicateurs de QoE, réalisée pour évaluer l'impact de la variation de la profondeur binoculaire perçue sur la QoE en TV S-3D. Pour cette expérimentation, des contenus synthétiques et naturels sont générés avec différents niveaux de profondeur binoculaire perçue, en contrôlant précisément les paramètres de prise de vue. Une limite concernant la plage de profondeur perçue maximale est proposée.

La partie III, composée des chapitres 7, 8 et 9, présente l'évaluation de l'impact d'autres problèmes techniques importants sur la QoE en TV S-3D: compression, format de représentation des images et asymétries entre vues. Le chapitre 7 se focalise sur l'impact de la compression JPEG-2000 sur des images fixes stéréoscopiques. Cinq indicateurs sont employés dans l'évaluation subjective de la QoE. Le chapitre 8, composé de deux expérimentations, étudie l'impact des formats de représentation 3D sur la QoE en utilisant un écran TV S-3D entrelacé ligne. La première expérience s'intéresse à l'impact de la réduction de résolution de formats S-3D compatibles 2D (formats dits « frame-compatible »). La deuxième expérience est conçue pour comparer la QoE de différents formats S-3D compatibles 2D compressés à différents débits. Le chapitre 9 a pour objectif d'évaluer l'impact des asymétries entre vues sur la QoE en TV S-3D. Des seuils perceptuels sont mesurés et recommandés pour différents types d'asymétries.

R 1. Les défis liés à la QoE en TV S-3D

La qualité d'expérience (QoE) est une mesure subjective de l'expérience client pour un service donné. Dans le cas de la TV S-3D, il s'agit de la mesure subjective de l'expérience de l'observateur, obtenue avec des images stéréoscopiques présentées sur un écran S-3D. Comparé à la TV 2D conventionnelle, la TV S-3D peut fournir des informations additionnelles de profondeur: la disparité binoculaire. Ceci peut augmenter la perception de profondeur et améliorer la QoE. En attendant, la TV S-3D ne correspond toujours pas à une représentation parfaite d'une scène naturelle. La plupart du temps, regarder des images stéréoscopiques sur un écran TV S-3D n'est pas exactement identique à regarder une scène naturelle. Ces divergences peuvent provoquer des problèmes perceptuels, induisant une baisse de la QoE et même produire inconfort et fatigue visuels. De plus, des questions techniques relatives à une chaîne de diffusion TV S-3D moderne ont également une influence potentielle sur la QoE. Dans ce chapitre, nous présentons les défis liées à la QoE dans le cadre d'un service de TV S-3D.

R 1.1 Les fondements de la perception de la profondeur

Chez l'homme, la perception de la profondeur est la capacité à voir et comprendre le monde tridimensionnel. C'est une des principales fonctions de notre système visuel. Nos yeux n'ont à leur disposition que des images rétiniennes bidimensionnelles mais pas de troisième composante dédiée à la perception de la profondeur. Ainsi, c'est une interprétation de repères physiologiques qui conduit à une perception efficace. La perception de la profondeur est la combinaison des images rétiniennes de nos deux yeux permettant d'extraire la meilleure et la plus convaincante des informations sur les trois dimensions de notre monde. À proprement parler, les observateurs ne voient pas la profondeur mais des objets dans la profondeur, et ils ne voient pas l'espace mais des objets dans l'espace.

Les sources d'information de la profondeur, c’est-à-dire les repères de profondeur, peuvent être classées par catégorie dans quatre groupes : l'information picturale (par exemple l’occlusion, la taille relative, la densité relative, la hauteur dans le champ visuel), l’information dynamique (la parallaxe de mouvement et la perspective), l’information oculaire (la convergence et l’accommodation) et l’information stéréoscopique (la disparité binoculaire).

Repères de profondeur et TV S-3D : focus sur la disparité binoculaire

Comparée à TV 2D, le repère de profondeur le plus important apporté par la TV S-3D est la disparité binoculaire. Le cerveau exploite la disparité binoculaire pour extraire l'information de profondeur à partir des images rétiniennes bidimensionnelles (stereopsis), réalisant une meilleure discrimination en profondeur.

La sensibilité des repères de profondeur

La sensibilité des differents repères de profondeur dépend de la distance de visualisation. Dans l'article (E.Cutting and M.Vishton, 1995), les auteurs ont présentent les résultats et les discussions de leur étude sur sur la sensibilité des critères de profondeur à différentes distances de visualisation. Ils définissent trois types d’espaces, représentant différentes plages de profondeur: l’espace personnel (moins de 3 mètres), l’espace d’action (de 3 à 15 mètres) et l’espace lointain (plus de 15 mètres).

Pour l'information picturale, l’occlusion est classée comme le repère de profondeur le plus sensible dans chacun des trois espaces (espace personnel, espace d’action et espace lointain). Pour la taille relative, la densité relative, la hauteur dans le champ visuel et la perspective aérienne, leur sensibilité croît avec l'augmentation de la distance.

Pour l'information dynamique, la perspective de mouvement est classée au troisième rang des repères de profondeur les plus importants à distance proche (espace personnel). Cependant, sa sensibilité diminue avec l'augmentation de la distance.

Concernant l'information oculaire, pour les grandes distances (espace d'action et espace lointain) la convergence et l’accommodation sont classées comme les repères de profondeur les moins importants. Cependant, pour les distances proches (espace personnel), elles sont des repères de profondeur très sensibles pour indiquer la distance absolue des objets (Goodwin, 1995) c’est-à-dire la distance perçue des observateurs aux objets.

Pour l'information stéréoscopique, c’est-à-dire la disparité binoculaire, elle se produit quand l'objet est situé dans le champ de vision binoculaire. Elle est classée comme second repère au niveau de la sensibilité parmi les neuf repères de l’espace personnel. C'est un repère très important et efficace pour les distances proches. C'est également sur elle que repose le principe de la TV S-3D. En TV S-3D, puisque la distance de visualisation est habituellement située dans l'espace personnel (moins de 3 mètres), ajouter l'information de disparité binoculaire fournira un repère de profondeur important au système visuel humain, augmentant de ce fait la perception de la profondeur.

R 1.2 De la vision binoculaire au système vidéo stéréoscopique

Les images stéréoscopiques modernes sont capturées par deux caméras décalées horizontalement. Ainsi, le système stéréoscopique moderne le plus simple pour l'acquisition et la visualisation d'images consiste en:

- Pour l'acquisition, deux caméras vidéo utilisées pour remplacer les yeux gauche et droit. L'information de disparité binoculaire est représentée par une légère différence horizontale entre l'image gauche et l'image droite, c'est-à-dire la disparité image.
- Pour la visualisation, les images gauche et droite enregistrées à partir des caméras fournies respectivement à l'œil gauche et à l'œil droit. Généralement, la séparation des images gauche et droite peut être effectuée en utilisant une technique d'affichage stéréoscopique dédiée comme par exemple les anaglyphes, la polarisation, les obturateurs actifs, une barrière de parallaxe ou un réseau lenticulaire. Dans cette partie, l'information de disparité binoculaire est visualisée sur l'écran comme le grandissement de la disparité image, c'est-à-dire la disparité écran.

La Figure R- 2 décrit le principe du système vidéo stéréoscopique le plus simple.

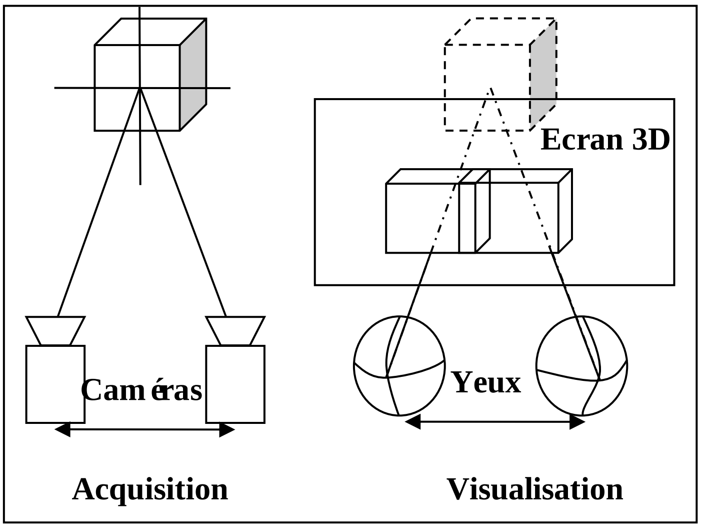


Figure R- 2: Le principe de base d'un système vidéo stéréoscopique.

Comparé à l'image 2D, les systèmes stéréoscopiques créent une illusion de sensation de profondeur en ajoutant l'information de disparité binoculaire. Cependant, il est important de noter que l'accommodation et la convergence ne sont pas reproduites comme dans le cas de la vision naturelle puisque le système de visualisation présente les images sur un écran plat. De plus, l'information de disparité binoculaire peut ne pas être identique en regardant la scène directement puisqu'elle dépend des paramètres d'acquisition et des paramètres de visualisation des images.

Des études (Lambooij et al., 2011, IJsselsteijn et al., 2000) ont confirmé que le sentiment d'immersion, de profondeur, d'aspect naturel et d'expérience visuelle augmentent sensiblement avec des images stéréoscopiques en comparaison avec la 2D.

D'autres études (Kooi and Toet, 2004, Woods et al., 1993, Yano et al., 2004, Yano et al., 2002, Lambooij et al., 2009a) ont montré que des problèmes tels que les asymétries entre les images droite et gauche, la déformation stéréoscopique, le découplage de l'accommodation et de la convergence peuvent également se produire. Ces effets négatifs peuvent provoquer un inconfort visuel et une fatigue visuelle comme détaillé dans la prochaine section.

R 1.3 L'impact de la TV S-3D sur l'inconfort visuel et la fatigue visuelle

Regarder des images stéréoscopiques est juste une illusion de profondeur accrue. Comparé à la visualisation d'images 2D, l'inconfort visuel et la fatigue visuelle sont plus fréquemment mentionnés en regardant des images stéréoscopiques. Les contradictions entre la visualisation d'images stéréoscopiques et celle de la scène réelle sont identifiées comme sources potentielles d'inconfort visuel et de fatigue visuelle.

Dans la littérature, l'inconfort visuel est employé de façon interchangeable avec la fatigue visuelle. Dans cette thèse, nous faisons une distinction entre ces concepts en proposant les définitions suivantes :

- **Inconfort visuel** : il est plus lié à un effet court terme qui peut être expliqué et mesuré subjectivement. En tant que tel, c'est un concept quelque peu ambigu, avec de nombreuses causes, des symptômes et des indicateurs associés. Par exemple, une gêne visuelle, des difficultés d'accommodation et de convergence ou encore un flou non naturel peuvent être mentionnés par les observateurs comme des symptômes et des indicateurs d'inconfort visuel.
- **Fatigue visuelle** : comme défini dans (Lambooij et al., 2009b), il s'agit d'une diminution de la performance du système visuel. C'est un critère objectivement mesurable dont la valeur mesure les processus d'adaptation à long terme du système visuel. Par exemple, une baisse de l'acuité visuelle et de l'acuité stéréoscopique, un dysfonctionnement dans la réponse de l'accommodation ou de la convergence ou encore une augmentation de la puissance de signaux EEG pourraient être utilisés comme indicateurs de fatigue visuelle.

L'inconfort visuel et la fatigue visuelle ne sont pas indépendants l'un de l'autre. La mesure subjective de l'inconfort visuel perçu fournit une indication sur la fatigue visuelle mesurable objectivement. L'accumulation de l'inconfort visuel à court terme aboutit à une fatigue visuelle. La non accumulation à court terme d'inconfort visuel pourrait seulement traduire une adaptation du système visuel et ne pas causer nécessairement une fatigue visuelle. De plus, une baisse momentanée dans les performances du système visuel n'est pas toujours reliée à la fatigue visuelle. Le système visuel possède des degrés de flexibilité et est capable de s'adapter à des conditions de visualisation altérées. Pour distinguer une fatigue visuelle significative d'une adaptation fonctionnelle non problématique du système visuel, il est nécessaire d'effectuer une vérification croisée avec l'inconfort visuel perçu.

Lors de la visualisation de contenus stéréoscopiques sur des écrans TV S-3D, divers facteurs peuvent engendrer inconfort visuel et fatigue visuelle tels qu'une disparité excessive sur l'écran, la non-corrélation de l'accommodation et de la convergence, les déformations du relief, l'asymétrie entre vues et les anomalies stéréoscopiques ou encore la violation de fenêtre.

Pour éviter inconfort visuel et fatigue visuelle, les images stéréoscopiques devraient remplir les conditions ci-dessous:

- Etre présentées dans la zone de confort de visualisation en profondeur pour éviter des disparités écran excessives et le découplage de l'accommodation et de la convergence ;
- Adapter les paramètres des caméras en considérant l'environnement de visualisation pour éviter les déformations stéréoscopiques ;
- Eviter les asymétries d'image ou, au moins, garantir le niveau des asymétries d'image au-dessous du seuil perceptuel ;
- Présenter les images dans des conditions non-dégradées pour éviter les anomalies stéréoscopiques ;
- Concevoir le contenu de la scène ou le rendu final en profondeur pour respecter la vision humaine.

R 1.4 Questions liées à la QoE dans une architecture de diffusion TV S-3D moderne

La production de contenus

Comparée à la production d'images 2D, la production d'images 3D stéréoscopiques exige une information additionnelle, c'est-à-dire l'information de profondeur binoculaire. Il existe différents systèmes de production 3D pour capturer et produire des images 3D stéréoscopiques comme par exemple les systèmes monoscopiques avec conversion 2D vers 3D (conversion automatique ou semi-automatique), les systèmes monoscopiques avec capteurs de profondeur additionnels, les systèmes stéréoscopiques traditionnels composés de 2 caméras (système avec miroir semi-transparent, rig à plat, etc.), les systèmes de multi-vues (plus de deux caméras, avec ou sans capteur de profondeur additionnel) et la production de contenus synthétiques.

En ce qui concerne les questions liées à la QoE des systèmes de production stéréoscopiques mentionnés ci-dessus, plusieurs remarques peuvent être faites :

- Pour les systèmes basés carte de profondeur, tel que les conversions 2D vers 3D et les systèmes monoscopiques avec capteur de profondeur additionnel, le manque d'information sur les couches d'occlusion et la précision de la carte en profondeur peuvent affecter la qualité finale des images stéréoscopiques reconstruites ;
- Pour les systèmes stéréoscopiques composés de deux caméras, la configuration caméra et les paramètres de prise de vue peuvent affecter la perception finale de la profondeur. De plus, la calibration des caméras est très importante pour éviter des asymétries d'image ;

- Les systèmes multi-vues semblent pouvoir reconstruire l'information 3D la plus précise. Cependant, la calibration des caméras est bien plus complexe ;
- La production graphique 3D peut fournir une excellente qualité pour chaque vue. Cependant, elle ne peut être employée que pour les scènes synthétiques.

Le format de la représentation 3D

Il existe divers formats de représentation 3D (Macchiarella, 2010, Gautier et al., 2010) disponibles dans la littérature, tels que le format vidéo stéréoscopique conventionnel, le format 2D plus profondeur, le format vidéo multi-vues (MVV) et le format vidéo multi-vues plus profondeur (MVD), le format vidéo 2D plus profondeur et données d'occlusion (LDV) et le format vidéo stéréoscopique avec carte de profondeur (DES). Dans cette section, les avantages de chaque format seront discutés ainsi que leurs inconvénients et limitations.

Les questions relatives à la qualité des formats de représentation 3D sont :

- Pour les formats dits "frame compatible", l'effet de la réduction de résolution peut affecter la qualité et exige des recherches supplémentaires ;
- Pour les formats basés carte de profondeur, même pour le format LDV, la qualité de reconstruction de la nouvelle vue n'est toujours pas comparable aux vues stéréoscopiques natives (Kerbiriou et al., 2010).

Le codage et la transmission

Les questions liées à la QoE en codage et transmission sont les suivantes :

- Pour le codage et la transmission des formats vidéo 3D compatibles 2D ("frame compatible"), ceux-ci peuvent nécessiter des débits plus élevés que les canaux 2D HD conventionnels (2D-HD) afin d'assurer la même qualité au niveau du rendu des textures que les images 2D-HD. Jusqu'ici, aucun débit optimal n'a été proposé.
- Pour les formats 3D pleine définition ("full definition") ou les formats de représentation 3D avancés tels que le format 2D plus profondeur ou le format LDV, ils augmentent la complexité calculatoire pour le codage et exigent un débit plus élevé pour la transmission.
- Le codage avec résolution "mixte" peut induire des asymétries entre vues avec un impact potentiel sur l'inconfort visuel et la fatigue visuelle.

Le terminal de visualisation

Nous classons les terminaux de visualisation TV S-3D selon s'ils nécessitent ou non l'utilisation de lunettes pour la séparation des vues gauche et droite. Les techniques nécessitant le port de lunettes sont les anaglyphes, la polarisation, les systèmes à obturation, les solutions de type « eyewear » et « Head Mounted Display » (HMD). Les techniques ne nécessitant pas le port de lunettes dédiées sont basées sur les barrières de parallaxe et les réseaux lenticulaires, qu'ils soient multi-vues ou avec système de suivi du regard.

La principale question liée à la QoE pour les terminaux de visualisation est qu'il n'existe pas d'écran dit "transparent". Asymétries colorées, perte de luminance, images fantômes, perte de résolution spatiale ou de résolution temporelle, contraintes de positionnement de l'observateur, etc. peuvent exister avec les différentes techniques et affecter la QoE.

R 1.5 Conclusion

Dans ce chapitre, nous avons discuté des principaux défis liés à l'évaluation de la QoE en TV S-3D. A partir des fondements sur la perception de la profondeur et de son impact potentiel sur la QoE en TV S-3D, quelques conclusions peuvent être tirées:

- Les téléspectateurs peuvent extraire l'information de profondeur à partir de neuf sources d'information différentes. La disparité binoculaire est l'une d'entre elles et elle est particulièrement sensible dans l'espace personnel (moins de 10 mètres de profondeur). La TV S-3D peut représenter la disparité binoculaire dans l'image. Ainsi, notre système de vision peut tirer profit de ce repère additionnel de profondeur pour produire une illusion de profondeur augmentée.
- Les systèmes stéréoscopiques ne produisent pas une représentation parfaite du monde 3D réel. Des divergences entre la visualisation TV S-3D et la visualisation de scènes réelles ont un impact potentiel sur la QoE. Si ces divergences sont assez importantes, elles provoquent de l'inconfort visuel ou même de la fatigue visuelle.

Ainsi, les méthodes d'évaluation de la QoE devraient permettre la mise en avant des avantages mais également les problèmes de la QoE en TV S-3D.

De l'examen des questions liées à la QoE sur la chaîne de diffusion TV S-3D, nous concluons :

- Il n'y a aucune solution parfaite proposée pour la diffusion TV S-3D qui peut fournir une résolution suffisamment haute à chaque œil sans exhiber des asymétries entre vues. Diverses techniques sont disponibles et leur impact sur la QoE n'est pas identique.

R 2. Les méthodologies pour évaluer la QoE 3D

Comme présenté dans le chapitre 1, diverses questions relatives à la QoE existent concernant les différentes techniques de la chaîne de diffusion TV S-3D. Elles ont un impact potentiel sur l'acceptation finale et le succès des services de TV S-3D. Ainsi, l'évaluation de la TV S-3D est urgente et importante pour beaucoup d'applications. Par exemple, elle peut être employée pour faciliter le processus de spécification pour les applications de bout en bout (par exemple, sélection des débits vidéo, de la technique d'affichage S-3D ou encore du codeur vidéo).

Dans le domaine scientifique et industriel, l'évaluation subjective est la manière la plus directe pour évaluer l'opinion des observateurs sur la QoE. L'évaluation subjective conventionnelle se concentre principalement sur l'évaluation de la qualité d'image. Mais concernant la QoE de la TV S-3D, la qualité d'image pourrait ne pas être un terme suffisant pour représenter la QoE. Par exemple, il ne peut pas directement mettre en avant les avantages (perception augmentée de la profondeur, etc.) et les problèmes (inconfort visuel, etc.) de la TV S-3D. De plus, concernant les spécifications actuelles des méthodes d'évaluation subjective de qualité telles que l'environnement de visualisation, elles ne sont pas adaptées aux nouvelles caractéristiques de la TV S-3D.

R 2.1 État de l'art : l'évaluation subjective de la QoE pour la TV S-3D

Les recommandations de l'UIT

Les méthodes subjectives normalisées pour l'évaluation de la qualité ont une longue histoire. La recommandation UIT-R BT.500-11- « méthodologie pour l'évaluation subjective de la qualité de la télévision », a été publiée en 2002 et reste la recommandation la plus réputée et la plus couramment employée dans le domaine de l'évaluation de la qualité des images. Pour évaluer les images en télévision stéréoscopique, la recommandation UIT-R BT.1438- « évaluation subjective des images stéréoscopiques en télévision » a été publiée en 2000 par l'UIT.

Cependant, suite au développement rapide des diverses techniques 3D stéréoscopiques ces dernières années, la recommandation UIT-R BT.1438 nécessite des spécifications supplémentaires prenant en compte les nouvelles caractéristiques de la TV S-3D. Ainsi, les groupes UIT-R WP6 et UIT-T SG9 ont adressé de nouvelles questions (Q.2 et Q.12 à l'UIT-T) et progressent dans la rédaction de nouvelles recommandations (ébauche). En attendant, le groupe d'experts de la qualité vidéo (VQEG), contributeur actif pour la plupart des questions de l'UIT-T SG9, a établi un nouveau projet « TV 3D » visant à étudier comment évaluer la qualité vidéo subjective en TV 3D.

Les études exploratoires

En parallèle des activités de normalisation internationale, dans la dernière décennie, beaucoup d'études exploratoires ont été réalisées pour une meilleure compréhension et une meilleure évaluation de la QoE des images stéréoscopiques. En passant en revue les études exploratoires destinées à évaluer la QoE en TV S-3D, les trois principales conclusions sont :

- Beaucoup d'études ont employé différents indicateurs de la QoE, ou des attributs subjectifs pour l'évaluation de la QoE des images stéréoscopiques comprenant la quantité de profondeur, la qualité de la profondeur, la qualité des textures et le piqué de l'image, le confort visuel (inconfort visuel, gêne visuelle et fatigue oculaire), la fatigue visuelle, l'expérience de visualisation (qualité globale de l'image ou l'expérience visuelle), l'aspect naturel, la présence et le plaisir. Il est nécessaire de clarifier qu'il n'y a aucune définition commune pour certains indicateurs de la QoE. Par exemple, l'indicateur " profondeur" peut se rapporter à la quantité de profondeur ou à la qualité de la profondeur. La qualité d'image peut se rapporter à la qualité des textures ou à la qualité globale de l'image. Ainsi, il peut être difficile de faire une comparaison équitable entre les études. Cependant, le point commun des études exploratoires destinées à évaluer la TV S-3D est que le concept conventionnel de « qualité » est trop limité pour représenter la QoE de la TV S-3D, et que des indicateurs multidimensionnels de QoE sont nécessaires.

- Parmi les études réalisées, les caractéristiques des tests subjectifs différaient. Par exemple, au sujet des conditions de visualisation, divers types et taille d'affichage d'écrans TV S-3D ont été employés sans spécifier le processus de calibration et la luminance. La règle pour déterminer la distance de visualisation était différente en fonction des études. La description des séquences source manquait parfois de précision (manque de spécification). La plupart des études n'a pas suivi les recommandations UIT-R BT.500 et UIT-R BT.1438, et ceci peut s'expliquer parce que les conditions générales de visualisation proposées par la recommandation UIT-R BT.500 ne sont pas adaptées aux applications 3D. Cela peut également induire des difficultés pour la comparaison des résultats des études.
- Pour la mesure de la fatigue visuelle, il n'y a pas encore de méthodes communes d'évaluation.

L'élaboration d'une nouvelle méthode d'évaluation subjective normalisée de la QoE devrait considérer les trois problèmes ci-dessus pour fournir des spécifications destinées à guider l'évaluation subjective afin d'obtenir des résultats subjectifs fiables, comparables et reproductibles.

R 2.2 Vers l'adaptation complète de l'évaluation subjective de la QoE en TV S-3D

Les méthodologies conventionnelles destinées à l'évaluation subjective de la qualité doivent s'adapter pour évaluer au mieux la QoE en TV S-3D. Comme la QoE est multidimensionnelle pour la TV S-3D, des indicateurs multidimensionnels de la QoE sont indispensables pour la représenter. De plus, la spécification des caractéristiques communes pour évaluer les images en TV S-3D est requise pour considérer les nouveaux facteurs de la TV S-3D puisqu'ils pourraient avoir des impacts potentiels sur la QoE.

Proposition d'indicateurs de la QoE

Le concept traditionnel pour évaluer la QoE, c'est-à-dire l'évaluation de la qualité d'image, n'est pas suffisante pour mettre en avant les avantages et les inconvénients des images stéréoscopiques. Par exemple, la qualité d'image n'est pas sensible à la profondeur perçue et aux problèmes de confort visuel. Au regard de la littérature, une des explications avancée pour mieux comprendre la QoE en TV-S 3D est que son évaluation devrait être multidimensionnelle. En récapitulant les propositions de la littérature, nous proposons d'employer les indicateurs ci-dessous pour évaluer la QoE en TV S-3D :

- **Qualité d'image 2D** : c'est la qualité du rendu de la texture. Dans le cas des images 2D, la qualité d'image 2D est identique à la « qualité d'image » traditionnelle. Cependant, dans le cas des images 3D, la qualité d'image 2D représente le jugement de la qualité de la texture à l'exclusion de l'information de profondeur et de la qualité de la profondeur.
- **Quantité de profondeur** : c'est la quantité de la profondeur perçue incluant la combinaison des repères monoculaires et binoculaires de profondeur.
- **Confort visuel** : l'inconfort visuel est lié à de multiples symptômes, par exemple une fatigue oculaire, les yeux secs et une vision double. La variation du confort

visuel peut être perçue comme, par exemple, la sensation d'une dégradation visuelle aussi bien qu'une sensation de difficultés de vision.

- **Rendu de la profondeur** : c'est la qualité de la profondeur perçue. Elle dépend de la préférence des observateurs sur des critères de base liés à l'étirement ou à la compression de la réalité et à la forme des objets.
- **Aspect naturel** : il représente l'évaluation de l'aspect normal des images, c'est-à-dire si la scène est plus ou moins représentative de la réalité.
- **Expérience visuelle** : c'est la qualité d'expérience globale (QoE) des images en termes d'immersion et qualité perçue globale.

A partir de la définition des six indicateurs de la QoE présentés ci-dessus, nous pouvons séparer ces indicateurs selon deux niveaux. Les indicateurs de plus haut niveau tels que l'expérience visuelle, l'aspect naturel et le rendu de la profondeur peuvent être une combinaison complexe de différentes décisions cognitives et perceptuelles. Les indicateurs de plus bas niveau tels que la qualité d'image, la quantité de profondeur et le confort visuel représentent les indicateurs de base de la QoE. Ils peuvent avoir un lien direct avec les paramètres techniques.

Les indicateurs ci-dessus visent à évaluer l'opinion à court terme ou instantanée de la QoE des images stéréoscopiques. D'autre part, la visualisation à long terme de la TV S-3D pourrait provoquer de la fatigue visuelle et influencer la QoE en TV S-3D, comme présenté dans le chapitre 1. Ainsi, la fatigue visuelle peut être employée comme indicateur à long terme de la QoE et être définie comme ci-dessous :

- **Fatigue visuelle** : c'est une diminution de la performance du système visuel. C'est un critère mesurable de manière objective, destiné à établir les processus adaptatifs à long terme du système visuel.

Les nouveaux facteurs affectant l'évaluation de la QoE en TV S-3D

En ce qui concerne l'évaluation subjective de la qualité vidéo, les critères communs comme décrit dans la recommandation UIT-R BT.500 ne prennent pas en considération les nouvelles caractéristiques de la TV S-3D. Ainsi, l'adaptation des méthodologies conventionnelles est nécessaire en considérant les nouveaux facteurs de la TV S-3D.

Les nouveaux facteurs affectant l'évaluation subjective de la TV S-3D sont récapitulés dans le Figure R- 1. La plupart de ces nouveaux facteurs exigent des tests supplémentaires pour recommander des spécifications précises. Dans cette thèse, plusieurs études sont destinées à contribuer au processus de spécification pour développer une nouvelle méthode subjective d'évaluation de la QoE pour la TV S-3D.

Tableau R- 1: Nouveaux facteurs influençant l'évaluation subjective en TV S-3D

Caractéristique	Facteur	Nouveau facteur
Conditions de visualisation générales	**Luminance et contraste**	Réduction de la luminance provoquée par un instrument optique additionnel, luminance minimum nécessaire pour maintenir la DOF (Depth of focus) , impact du « crosstalk » sur le contraste
	Eclairage de la salle et de l'arrière-plan	Distance minimum nécessaire entre l'écran et l'arrière-plan, technologie d'éclairage de la salle
	Résolution du moniteur	Valeurs minimales pour les résolutions temporelles et spatiales par vue, résolution stéréoscopique
	Distance de visualisation	Distance de visualisation préconisée (DVD) fixée par le fabricant de l'écran, distance de visualisation préférée (PVD)
	Position de visualisation	Limitation des déformations géométriques 3D, réduction de la luminance, position de de visualisation optimale pour les écrans autostéréoscopiques
	Rendu de la profondeur	Limites supérieures pour DOF et la disparité binoculaire
Signaux source	**Format vidéo**	Besoins pour les formats de représentation de la profondeur
	Vidéo format conversion	Spécifications de la précision des conversions
Sélection des séquences de test	**Complexité du contenu vidéo**	Outils de mesure de la complexité en profondeur des contenus
Observateurs	**Nombre**	Re-évaluation nécessaire pour garantir la stabilité et la fiabilité des résultats
	Performance de la « stereopsis »	Mesure de la "stereopsis", précision, différences oculaires, etc.
Session de test	**Durée de visualisation**	Re-évaluation de la durée de présentation, vote, durée d'une session
Analyse des résultats de test	**Facteur "observateur"**	Critère de rejet, détection de distributions bimodales
	Analyse des critères multidimensionnels	Méthodes statistiques pour l'analyse, par exemple relation, interaction et combinaison des indicateurs de QoE mesurés subjectivement
Méthode de test	**Fatigue visuelle**	Mesure objective de la fatigue visuelle
	Indicateur subjectif de la QoE	Indicateurs de la QoE multidimensionnelle

R 2.3 Conclusion

Dans ce chapitre, à partir de l'examen des protocoles d'évaluation de la QoE, plusieurs conclusions peuvent être mentionnées :

- Les méthodes conventionnelles d'évaluation subjective de la qualité ne sont pas suffisantes pour évaluer la qualité des images stéréoscopiques. L'UIT et VQEG travaillent sur de nouvelles méthodes d'évaluation subjective de la qualité pour les images stéréoscopiques.
- Pour les études exploratoires, divers indicateurs de QoE ont été employés. Cependant, il n'existe aucune définition commune de ces indicateurs de QoE. De plus, l'environnement et les conditions de visualisation changent entre les études. Et, il est donc difficile de comparer les résultats des différentes études.

Les contributions principales de ce chapitre sont :

- La proposition et la définition d'indicateurs multidimensionnels de QoE : la qualité d'image 2D, la quantité de profondeur, le confort visuel, le rendu de la profondeur, l'aspect naturel et l'expérience visuelle ainsi que la fatigue visuelle.
- Une discussion focalisée sur l'évolution des méthodes d'évaluation subjective de la QoE en TV S-3D : De nouveaux facteurs sont proposés afin de développer une nouvelle méthode d'évaluation de la QoE en TV S-3D. Cette contribution est destinée à juger de l'intérêt de cette nouvelle méthode ainsi qu'à contribuer à la définition de la nouvelle méthodologie d'évaluation subjective de QoE des images stéréoscopiques pour la TV S-3D.

Les propositions ci-dessus seront employées dans les études suivantes de cette thèse.

R 3. Caractérisation des écrans TV S-3D

Comme présenté au chapitre 1, il n'existe actuellement aucun affichage "transparent" disponible, ne posant pas de problème pour l'évaluation de la QoE. Ainsi, la caractérisation de l'affichage TV S-3D est essentielle afin de choisir l'affichage le mieux adapté aux expérimentations ou d'adapter la performance de l'écran de manière optimale. La performance de l'écran devrait également être considérée pour analyser les résultats expérimentaux des évaluations sur la QoE. Cependant, les méthodes subjectives conventionnelles d'évaluation de la qualité telle que la recommandation UIT-R BT.500 souffrent d'un manque de spécifications sur les écrans d'affichage TV S-3D. Dans cette section, nous nous concentrerons sur deux des plus importants facteurs permettant de caractériser les écrans TV S-3D : le rendu de la luminance et le rendu de la profondeur.

R 3.1 Le rendu de la luminance

En TV S-3D, le rendu de la luminance couvre deux types de caractéristiques présentées ci-dessous :

- Caractéristiques 2D : nécessite des mesures sur une seule vue;
- Caractéristiques 3D : exige des mesures sur plus d'une vue.

De plus, il est également important de distinguer les modes 2D et 3D dans les écrans TV 3D. La plupart des techniques courantes d'affichage TV 3D sont des versions étendues ou avancées d'écrans 2D avec la fonctionnalité de séparer et de fournir des vues différentes aux yeux gauche et droit.

Ainsi, pour mesurer le mode 2D d'un écran stéréoscopique, il ne faut mesurer que les caractéristiques 2D.
Pour mesurer le mode 3D d'un écran stéréoscopique, il faut considérer les deux : les caractéristiques 2D par vue et les caractéristiques 3D. Les caractéristiques 2D du rendu de la luminance pour les écrans TV S-3D incluent : la fonction de transfert de la luminance (fonction gamma), le gamut de couleur, la résolution, la performance temporelle, le temps de réponse, l'uniformité et l'angle de vue. Les caractéristiques 3D du rendu de la luminance pour les écrans TV 3D incluent les images fantôme et la position de visualisation.
A partir des caractéristiques 2D et 3D ci-dessus, d'autres caractéristiques des écrans TV S-3D sont exigées pour évaluation, comme par exemple le format d'image et les traitements d'images intégrés dans l'écran comme par exemple l'affichage des images entrelacées.

Etude de cas

Une étude destinée à mesurer les performances de quatre écrans TV 3D différents est présentée dans cette thèse.

Les principaux objectifs de ces mesures sont :

- Justifier que les performances des différents écrans TV 3D peuvent être diverses et variées. Ainsi, la mesure et l'ajustement des écrans TV 3D ou le choix de l'affichage optimal pour les évaluations subjectives de qualité est indispensable.
- Sélectionner le meilleur écran en termes de performance d'affichage pour l'utiliser comme plateforme de visualisation pour les études de cette thèse.

Plusieurs conclusions remarquables de ces études sont résumées ci-dessous :

- La réduction de luminance pour les écrans basés LCD plus filtre polarisé passif est d'environ 50-60% ; En revanche, la réduction de luminance se situe autour de 70% pour les écrans basés PDP plus obturateur actif (active shutter).
- Deux des écrans mesurés dans l'étude de cas ont une mauvaise fonction de transfert en luminance, c'est-à-dire que la valeur gamma n'est pas égale à la valeur normalisée 2,2. Ceci peut augmenter la visibilité des artéfacts de codage comme par exemple l'effet de bloc et les effets de quantification de la luminance.
- La mesure du temps de réponse a été obtenue par la mesure directe du temps de montée (noir au blanc) et du temps de descente (blanc au noir) par un dispositif de la société ELDIM. Cependant, cette méthode peut ne pas être appropriée pour identifier les performances temporelles des écrans PDP puisque le principe de rafraichissement temporel est différent de l'affichage des écrans à cristaux liquides.
- En ce qui concerne la valeur des images fantômes (crosstalk), il a été mesuré en envoyant une image blanche sur la vue gauche et une image noire sur la vue droite. Le niveau de « crosstalk » est alors le rapport entre le niveau de luminance mesuré sur la vue droite à travers les lunettes et le niveau de luminance de la vue gauche également mesuré à travers les lunettes. Cependant, la perception du « crosstalk » dépend également de la luminance, du contraste et de la disparité .

Dans cette thèse, la sélection de l'écran destiné aux tests subjectifs est basée sur les spécifications ci-dessous :

- La valeur gamma pour la fonction de transfert en luminance est proche de la valeur nominale 2.2 afin de reproduire des niveaux de luminance corrects.

- La luminance est supérieure à 200 cd/m2 en mode 2D et supérieure à 100 cd/m2 en mode 3D pour les 2 écrans retenus.
- Le temps de réponse blanc-noir-blanc varie de 8 à 12 ms.
- Le niveau de « crosstalk » est inférieur à 5%.

R 3.2 Le rendu de la profondeur

En comparaison avec les écrans 2D, l'élément clé des écrans TV S-3D réside dans sa capacité à restituer l'information de disparité binoculaire aux téléspectateurs afin d'augmenter la perception de la profondeur. Le rendu de la profondeur des écrans TV S-3D est lié à l'environnement de visualisation (par exemple, la distance de visualisation), aux propriétés d'affichage des écrans TV S-3D (par exemple, la taille des pixels et la taille de l'écran) et aux contraintes liées au système visuel humain (par exemple, la profondeur de mise au point de l'œil). La façon de représenter la capacité de rendu de la profondeur des écrans TV S-3D demeure toujours une question en suspens. Dans cette section, nous présentons un modèle théorique pour analyser le rendu de la profondeur des systèmes TV S-3D. En combinant les paramètres physiques et les contraintes perceptuelles, nous définissons la capacité de rendu de la profondeur ainsi que l'intervalle angulaire des plans de profondeur. A partir des définitions proposées, la capacité de rendu de la profondeur sera analysée et discutée pour différents types d'affichages stéréoscopiques.

Modélisation du rendu de la profondeur des systèmes TV S-3D

Nous établissons un modèle géométrique pour représenter la capacité à rendre la profondeur en combinant les paramètres physiques de l'environnement de visualisation avec les contraintes de la DOF et de la disparité binoculaire.

Les paramètres physiques sont : l'écart interpupillaire, la distance de visualisation, la résolution, le voxel stéréoscopique, le plan de profondeur.

Les contraintes perceptuelles sont la profondeur de mise au point de l'œil (DOF pour Depth of Focus) et la valeur limite de la disparité binoculaire. Ces données perceptuelles permettent de définir la zone de confort visuel et la disparité non croisée maximale en pixel.

Pour combiner les paramètres physiques et les contraintes perceptuelles, nous définissons les facteurs ci-dessous pour représenter la capacité de rendu de la profondeur de la TV S-3D :

- ***Capacité de rendu en profondeur en pixel*** : la capacité de rendu en profondeur est définie comme le nombre de plans en profondeur qui peuvent être représentés dans une zone de visualisation confortable autour de l'écran.
- ***Intervalle angulaire de plan de profondeur*** : l'intervalle angulaire de plan de profondeur est la distance entre deux plans de profondeur adjacents, fournissant une mesure de la quantification en profondeur. La valeur reste presque constante si elle est mesurée en unités angulaires au lieu d'unités mètriques.

Analyse des capacités de rendu du relief de différents écrans TV S-3D

Quatre types d'écrans TV S-3D ont été analysés incluant des affichages pleine définition, avec entrelacement des lignes ou entrelacement des colonnes ainsi que des affichages autostéréoscopiques, en tenant compte de la taille des écrans (ordinateur de bureau, TV et cinéma). Les discussions et conclusions de cette étude sont récapitulées ci-dessous :

- La capacité à rendre la profondeur dépend principalement de deux paramètres : la distance de visualisation et les propriétés de l'écran. Dans notre comparaison, la meilleure solution est le système basé sur deux projecteurs HD avec la pleine résolution. Ce système fournit une région de confort visuel raisonnablement bonne et assez de plans en profondeur afin de donner une bonne perception de la profondeur à l'observateur. Il comporte également un champ visuel de 30° qui est nécessaire pour créer une sensation remarquable de réalité. Il peut être considéré comme le système de référence avec une capacité de rendu de la profondeur optimale.
- Pour les écrans de petite taille, par exemple les écrans d'ordinateur de bureau avec la pleine résolution, ou les TV avec affichage entrelacé ligne, une plus longue distance de visualisation pourrait être prioritaire sur le champ visuel afin de garantir une zone de confort de visualisation plus large. La zone de confort de visualisation, la capacité de rendu de la profondeur et l'intervalle angulaire de plan en profondeur sont des fonctions de la distance de visualisation. Ainsi, une augmentation de la distance de visualisation de façon appropriée peut augmenter la zone de confort autour du plan de l'écran, permettre la fusion de disparités plus grandes et diminuer l'intervalle angulaire des plans en profondeur. L'inconvénient est la réduction de champ visuel. De même, pour les écrans multivues, l'augmentation de la distance de visualisation contribuera non seulement à une zone de visualisation confortable mais également à une réduction des artéfacts due à la quantification de la profondeur.
- En plus de la capacité à rendre la profondeur, la disparité du contenu affecte également le rendu de profondeur. Pour la production stéréoscopique, les vues gauche et droite sont souvent enregistrées et stockées dans un format stéréoscopique conventionnel. Dans ce cas, l'étendue des disparités du contenu est fixe et ne peut pas être modifiée sans traitement considérable et avec perte. Pour chaque écran, la capacité de rendu en profondeur de chaque écran est fournie comme limite supérieure du confort de visualisation. Quand l'étendue des disparités du contenu est en dehors de l'étendue préconisée pour chaque écran, les observateurs pourraient être incapables de fusionner les images. A l'opposé, quand l'étendue des disparités du contenu est beaucoup plus petite que la capacité de rendre la profondeur, les téléspectateurs pourraient percevoir un effet de profondeur appauvri. Comme la capacité de rendre la profondeur couvre une étendue allant de 22 pixels pour le cinéma numérique à 82 pixels pour une solution de projection TV HD, il pourrait être difficile d'employer le même contenu dans une expérience subjective menée avec différents types d'écrans. En termes d'évaluation subjective de la qualité vidéo, le choix des séquences de test devrait respecter le principe que la disparité du contenu devrait être adaptée à la capacité du rendu de la profondeur de l'écran. De plus, l'analyse ou la comparaison des résultats d'évaluations subjectives devrait également considérer soigneusement ces deux facteurs. Quand la disparité du contenu est plus grande que la capacité du

rendu la profondeur de l'écran, les observateurs peuvent avoir des difficultés à fusionner des images.

R 3.3 Conclusion

Dans ce chapitre, nous nous sommes attachés à proposer une caractérisation des écrans TV S-3D à partir du rendu de la luminance et du rendu de la profondeur.

Pour le rendu de la luminance, de nouvelles caractéristiques ont été présentées et discutées. Une étude de cas comparant quatre écrans TV S-3D différents a été réalisée afin de mettre en évidence les différences de rendu de la luminance entre les différents écrans TV S-3D.

Pour le rendu de la profondeur, nous avons défini de nouveaux facteurs pour représenter la capacité du rendu de la profondeur des écrans TV S-3D, en considérant les paramètres physiques et les contraintes perceptuelles. A partir de ces facteurs et des définitions proposées, différents écrans TV S-3D ont été analysés. Les résultats indiquent que la capacité de rendre la profondeur des écrans TV S-3D dépend principalement de la distance de visualisation, de la taille des pixels, de la taille d'écran et de l'organisation des pixels de l'écran pour les vues droite et gauche.

Ainsi, la caractérisation des écrans TV S-3D est nécessaire puisque le rendu de la luminance et le rendu de la profondeur de la TV S-3D dépend de divers facteurs qui peuvent affecter la QoE de la TV S-3D.

R 4. Mesure de la fatigue visuelle dans des conditions de visualisation optimales

La fatigue visuelle est une diminution des performances du système visuel. Lors de la visualisation d'images stéréoscopiques sur des écrans TV S-3D, deux hypothèses peuvent se présenter pour expliquer la fatigue visuelle :

- L'hypothèse pessimiste est que les techniques stéréoscopiques actuelles se sont développées sans prendre en compte le fonctionnement du système visuel humain. Par exemple, l'effet du découplage de l'accommodation et de la convergence est intrinsèque à ces systèmes. Ainsi, la fatigue visuelle est un problème inhérent et inévitable pour des techniques stéréoscopiques actuelles.
- L'hypothèse optimiste est que le système visuel humain est fait pour s'adapter facilement aux changements de fonctionnement. Dans ce cas, si les images stéréoscopiques sont présentées dans les limites acceptables du système visuel afin de ne pas induire ni accumuler d'inconfort visuel à court terme, la visualisation 3D stéréoscopique à long terme peut juste être une adaptation simple du système visuel qui ne devrait pas causer de fatigue visuelle.

De plus, la mesure de la fatigue visuelle n’est pas encore une question résolue. Afin de valider les hypothèses ci-dessus, et pour étudier les méthodes de mesure de la fatigue visuelle, un test de mesure de la fatigue visuelle dans des conditions de visualisation optimales a été réalisé. Les conditions de visualisation optimales de cette étude signifiaient la sélection et l'adaptation de l'environnement de visualisation, des stimuli expérimentaux et de l'écran TV S-3D afin d'éviter les problèmes d'inconfort visuel comme les disparités excessives, les asymétries d'image et les variations brutales des objets dans la profondeur. Comparé aux études précédentes, cette étude apporte les nouveautés suivantes :

- Deux contenus de sport, d'une heure chacun, l'un en 2D et l'autre en 3D, ont été employés comme stimulus.
- Les stimuli et les conditions de visualisation ont été choisis pour garantir strictement la profondeur perçue à l'intérieur de la zone de confort de visualisation c'est-à-dire ±0.2 dioptries. Les stimuli ne présentaient pas de mouvements saccadés dans l'axe de la profondeur.
- Les asymétries d'image ont été corrigées en postproduction.
- L'écran 3D stéréoscopique a été choisi et ajusté pour obtenir des performances optimales.
- La mesure de la fatigue visuelle a été réalisée par :
    - des méthodes objectives incluant un test de vision et des mesures électro-encéphalographiques (EEG)
    - une méthode subjective composée d'un questionnaire, rempli en début et en fin de test.

R 4.1 Méthodes objectives et subjectives

Dans cette expérience, trois types de test ont été conçus afin de détecter et mesurer la fatigue visuelle objectivement et subjectivement : Tout d'abord, le test de vision, réalisé avant et après une session de visualisation d'un contenu d'une heure, ensuite, le questionnaire, rempli avant et après la session de visualisation du contenu d'une heure, enfin, la mesure continue des signaux EEG sur 16 canaux pendant la session de visualisation vidéo d'une heure.

Le test de vision

Le test de vision a été réalisé à l'aide de l'équipement ERGOVISION d'Essilor qui est conçu pour aider le chercheur à explorer la fonction visuelle. Six tests préréglés liés à la vision binoculaire ont été choisis comme indicateurs de fatigue visuelle : phories, fusion, acuité monoculaire des yeux droit et gauche (acuité visuelle), fatigue visuelle, acuité stéréoscopique.

Le questionnaire

Deux questionnaires, c'est-à-dire un questionnaire rempli avant et un autre après la session de visualisation d'une heure ont été conçus pour évaluer la fatigue visuelle.

Pour le questionnaire rempli avant visualisation, cinq points principaux ont été évalués :

- Problème général de santé ;
- Problème général de santé lié à la vision ;
- Symptôme général de fatigue ;

- Symptôme de fatigue visuelle, principalement lié à des symptômes oculaires constatés sur le moment ;
- Symptôme de fatigue visuelle, se rapportant à des symptômes oculaires constatés à l'issue de tâches ou d'activités.

Les observateurs étaient invités à remplir ce questionnaire avant le test pour rapporter leur état de santé et plus particulièrement celui lié à la vision.

Le questionnaire réalisé après la session de visualisation était composé de trois parties :

- Symptôme général de fatigue ;
- Symptôme de fatigue visuelle (sur le moment) directement lié à l'œil ;
- Symptôme de fatigue visuelle (activité).

Les mesures EEG

La solution « Active Two » de Biosemi (16 canaux) a été employée pour enregistrer l'activité du cerveau des observateurs pendant une session de visualisation d'une heure. Les positions des électrodes constituent un sous-ensemble des emplacements du système international 10-20. Les électrodes sont numérotées Fp1, Fp2, F4, Fz, F3, T7, C3, Cz, C4, T8, P4, Pz, P3, O1, Oz, O2, avec F pour Frontal, T pour Temporal, P pour Pariétal, O pour Occipital (correspondant aux lobes du cerveau) et C pour la ligne Centrale. Toutes les données des canaux ont été référencées par rapport au canal Cz lors de la phase de post-traitement.

R 4.2 Le déroulement du test

Le test a été conçu de la façon suivante :

1. Equipement : le test s'est déroulé dans une salle en conformité avec la recommandation de l'UIT-R BT. 500. Un écran stéréoscopique Panasonic LCD de 50 pouces à lunettes actives (120 Hz) a été utilisé comme écran de visualisation final. La distance de visualisation était fixée à 3,5 fois la hauteur de l'écran.
2. Observateurs : 9 observateurs non experts ont été recrutés pour participer à ce test.
3. Stimuli : des vidéos enregistrées lors du tournoi de tennis de Roland Garros ont été utilisées comme stimuli. Un des contenus était le final homme présenté dans des conditions 2D et le second contenu était la finale femme présentée dans des conditions 3D. La profondeur finale restituée des stimuli était inférieure à ±0,15 dioptries pour éviter toute disparité excessive. La post-production, réalisée par une société professionnelle, a été réalisée directement lors de l'étape d'acquisition pour éviter toute asymétrie entre vues.
4. Procédure : pour chaque testeur, deux sessions ont été conduites à des jours différents afin d'éviter les interactions entre chaque session, une pour la condition 2D et l'autre pour la condition 3D. L'ordre de passation des deux conditions a été inversé pour la moitié des testeurs (4 testeurs ont effectué la condition 2D en premier et les 5 autres testeurs ont effectué la condition 3D en premier). La condition de test (2D ou 3D) n'était pas communiquée au testeur avant le test. De plus, le testeur devait porter des lunettes pour les 2 conditions.

Pour chaque session, la procédure était la suivante :

1. Le sujet doit répondre au questionnaire « avant visualisation » ;
2. Les électrodes de l'EEG sont installées et calibrées selon le manuel du système « Active Two » de la société Biosemi ;
3. Le sujet regarde la vidéo pendant une heure tandis que le signal EEG est enregistré continuellement pendant la session entière ;
4. Quand la visualisation d'une heure est terminée, le test de vision est répété ;
5. Le sujet doit répondre au questionnaire « après test ». A noter que, le jour même du test, il est demandé à chaque sujet de contacter l'expérimentateur en cas de symptômes qu'il suspecte être relié au test.

R 4.3 Analyse des résultats

L'analyse du test de vision a montré que quatre testeurs n'avaient pas de changement de performance après une heure de visualisation vidéo dans des conditions 2D et 3D. Trois sujets ont eu de meilleures performances pour certains tests après une heure de visualisation, que ce soit pour les conditions 2D ou 3D. Ceci peut s'expliquer par l'effet d'entraînement. Seulement deux sujets ont eu une dégradation des performances dans les tests de phories et d'acuité stéréoscopique. Un test de Student par paires a été réalisé pour comparer les résultats du test de vision (en votant 0 pour « même performance », 1 pour « Performance meilleure », et -1 pour « Performance plus mauvaise ») entre les conditions 2D et 3D pour chaque critère de test. Cependant, aucune différence significative n'a été constatée.

D'un point de vue général, l'analyse du questionnaire montre qu'il n'y a pas d'évidence à ce que des conditions de visualisation 3D produisent plus de fatigue visuelle que celles de la 2D.

Concernant le test EEG, les résultats indiquent que dans le lobe Frontal, les plus hautes densités de puissance significatives pour les conditions 3D sont principalement situées dans la bande des fréquences béta pour 5 des canaux EEG, et dans la bande des fréquences Gamma pour 3 des canaux EEG (F3, Fz, F4). Dans le lobe Temporal et sur la ligne centrale, les courbes des spectres de puissance 2D et 3D sont assez similaires et il n'existe pas de différence statistiquement significative. Dans les lobes Occipital et Pariétal, les plus grandes densités de puissance significatives sont principalement situées dans la bande de fréquence Gamma pour tous les canaux. De plus, afin d'analyser les changements temporels des signaux EEG, chaque heure de données EEG a été segmentée de façon régulière en 3 parties : de 0 à 20 minutes, de 20 à 40 minutes et de 40 à 60 minutes. Il n'existe pas de différences significatives concernant les variations temporelles pour les conditions 2D et 3D pour la plupart des canaux. Cependant, en comparant les conditions 2D et 3D pour différentes périodes temporelles, il semble que la plage de bandes de fréquences significatives tend à augmenter dans la seconde période (de 20 à 40 minutes) mais diminue dans la troisième période (allant de 40 à 60 minutes)

Les principaux résultats de l'expérience sont récapitulés ci-dessous :

- A partir du test de vision et du questionnaire, il n'y a aucune évidence significative indiquant qu'une heure de visualisation 3D ait causé une fatigue visuelle de plus haut niveau qu'une heure de visualisation 2D.
- Pour la plupart des canaux de l'EEG situés dans le frontal, les lobes pariétaux et occipital, l'énergie des bandes bêta et/ ou gamma est plus haute pour la condition 3D que pour la condition 2D.
- En segmentant les données EEG d'une heure en trois parties (0-20 minutes, 20-40 minutes, 40-60 minutes), l'analyse des résultats des signaux d'EEG indique que le signal EEG dans les deux conditions 2D et 3D n'a montré aucun changement crucial à mesure que la durée de visionnement augmentait. Cependant, en comparant la 2D et le spectre de puissance 3D dans les mêmes périodes de temps, il semble que l'étendue de bande significative où la 3D a une densité de puissance plus élevée que la 2D, tend à augmenter dans la deuxième période (20 à 40 minutes) mais à réduire dans la troisième période (40 à 60 minutes).

En comparant ces résultats à ceux de la recherche précédente, Li et al. ont rapporté que :

- A partir du test subjectif, le niveau de fatigue visuelle en 3D était plus élevé qu'en 2D.
- A partir des mesures EEG, dans la plupart des canaux, la puissance des hautes fréquences (>12 hertz) était plus forte en 3D plutôt qu'en 2D et qu'elle tendait à augmenter à mesure que la durée de présentation augmentait.

La principale raison qui expliquerait la conclusion différente entre cette étude et l'étude précédente est peut être liée au contenu et à l'environnement de visualisation. Dans cette étude, l'environnement de visualisation et les contenus à visualiser ont été choisis pour garantir une visualisation confortable. Notre conclusion peut indiquer que la puissance plus élevée dans les bandes bêta et gamma n'est pas nécessairement liée à la fatigue visuelle. Elle peut également avoir un lien avec une concentration active quand les gens sont plus immergés dans les contenus 3D.

Ainsi, les conclusions tirées de cette étude sont :

- Si l'environnement de visualisation et le contenu peuvent être optimisés, la visualisation 3D ne génère pas de fatigue visuelle.
- En ce qui concerne l'activité cérébrale lors de visualisations 2D et 3D, il y a quelques différences significatives, particulièrement dans la puissance des bandes bêta et gamma, dans les parties frontale et postérieure du cerveau. Cependant, il peut ne pas y avoir de lien avec la fatigue visuelle.

R 5. Nouvelle proposition de règles de prise de vue stéréoscopiques pour optimiser la QoE en TV S-3D

Concernant la prise de vue stéréoscopique et le rendu de la profondeur, deux principaux facteurs sont supposés affecter la QoE en TV S-3D.

Le premier facteur est la déformation stéréoscopique dérivée de la relation géométrique qui existe entre l'espace des caméras et l'espace de visualisation. Elle indique la différence de géométrie existant entre la scène réelle et celle restituée lors de la visualisation des images stéréoscopiques. Pour éviter la déformation stéréoscopique, il est indispensable de prendre en considération les paramètres de la scène et les paramètres de visualisation lors du calcul des paramètres de prise de vue.

Le second facteur est la zone de confort de visualisation qui est définie comme la limite de la fusion binoculaire et de la profondeur de mise au point de l'œil. Pour cela, il faut éviter des disparités binoculaires excessives et la dissociation des fonctions de convergence et d'accommodation des yeux. Si l'objet en profondeur est en dehors de la zone de confort, la visualisation de cet objet peut induire un inconfort visuel, et de ce fait dégrader la QoE. Généralement, la zone de confort de visualisation est définie en prenant en considération l'environnement de visualisation final.

Ainsi, pour optimiser la QoE en TV S-3D, ces deux facteurs devraient être pris en compte. Les objectifs des études de cette section sont :

- Proposer des facteurs et des seuils pour définir la déformation stéréoscopique et la zone confortable de visualisation ;
- Proposer des règles pour déterminer les paramètres de prise de vue basés sur l'optimisation de la déformation stéréoscopique et de la zone de confort de visualisation ;
- Concevoir des tests subjectifs sur la QoE pour vérifier les règles de prise de vue proposées ;

R 5.1 Nouvelle proposition de règles de prise de vue stéréoscopiques basées sur la déformation stéréoscopique et la zone de confort de visualisation

En analysant la géométrie de l'espace des caméras et de l'espace de visualisation, la distorsion stéréoscopique des formes est définie comme la variation de la forme autour du plan de profondeur z :

$$\boldsymbol{D_s = \frac{Vb}{Bz + Mfb(1 - \frac{z}{d_{cov}})}}$$

Avec V pour la distance de visualisation, b pour l'entraxe des caméras (distance entre les caméras droite et gauche), M pour le facteur de grandissement (rapport entre la taille de l'écran et celle du capteur de la caméra), f pour la distance focale et d pour la distance de convergence. Quand D_S est égale à 1, la forme 3D autour du plan de profondeur z dans l'espace de visualisation est maintenue à la même valeur que dans l'espace caméra. Quand D_S n'est pas égale à 1, il existe une distorsion stéréoscopique de la forme. Par exemple, un cube peut être perçu comme un cuboïde et un objet rond comme un objet ovale en cas de distorsion stéréoscopique de la forme.

De plus, en combinant les propositions de la littérature relative au confort visuel en TV S-3D, la valeur la plus contraignante de 0.2 dioptrie est choisie dans cette thèse comme limite générale de la zone de confort de visualisation.

Pour optimiser la prise de vue stéréoscopique en TV S-3D en considérant l'effet de la déformation stéréoscopique et de la zone de confort de visualisation, trois règles de prise de vue sont proposées dans cette section.

Règle de prise de vue n°1

Adapter les paramètres de la scène ou de la caméra stéréoscopique pour proposer, autour de la région d'intérêt de la scène, un facteur de déformation des formes aussi proche que possible de un.

Règle de prise de vue n°2

Garantir le maintien de la profondeur de la scène (distance entre le premier et l'arrière-plan) dans la zone de confort de visualisation en adaptant les paramètres de la scène ou les paramètres de la caméra stéréoscopique.

Règle de prise de vue n°3

Quand les règles 1 et 2 ne peuvent pas être respectées simultanément, la règle 2 est prioritaire par rapport à la règle 1, ce qui signifie que le confort visuel est plus important que la déformation de la forme des objets.

R 5.2 Vérification des règles de prise de vue optimales proposées

Dans la section précédente, des règles de prise de vue stéréoscopiques ont été proposées pour éviter la déformation stéréoscopique et pour garantir une visualisation confortable. Cependant, toutes les propositions sont fondées sur des analyses ou des hypothèses théoriques. Leurs impacts perceptuels sur la QoE de la TV S-3D doivent être confirmées. Dans cette section, nous avons défini un test subjectif permettant d'étudier les impacts perceptuels des règles de prise de vue stéréoscopiques proposées.

La génération des images (synthétique)

Pour vérifier les règles de prise de vue optimales proposées, il est nécessaire de générer ou de filmer des contenus stéréoscopiques. Afin d'éviter les asymétries entre vues telles que les asymétries géométriques ou colorées, les contenus stéréoscopiques ont été générés avec des outils dédiés au graphisme 3D (contenus synthétiques). 5 scènes représentant différents niveaux de profondeur ont été générées à partir d'un projet d'animation open source appelé « Big buck bunny ». Pour chaque scène, 5 conditions différentes ont été définies afin de justifier les règles de prise de vue proposées :

- Condition 1 : image 2D. La vue gauche de la paire stéréoscopique a été utilisée telle quelle comme image 2D.
- Condition 2 : DOF égale à 0,1. Il s'agit de la valeur minimale pour assurer le confort de visualisation et permettant de générer une compression des formes stéréoscopiques ($D_S < 1$).

- Condition 3 : DOF égale à 0,2. Cette valeur est suggérée par la recommandation BT. 1438 de l'UIT-R (ITU, 2000) comme seuil de profondeur ou de netteté pour maintenir le confort visuel.
- Condition 3 : DOF égale à 0,3. Cette valeur est suggérée par (Lambooij et al., 2009a) comme seuil de profondeur ou de netteté pour maintenir le confort visuel.
- Condition 5 : D_s^{RoI} est égale à 1, c'est-à-dire qu'il n'existe pas de distorsion des formes à la distance du plan d'intérêt.

Pour les conditions 2 à 4, les différentes valeurs de DOF représentent différents niveaux de zone de confort de visualisation. Les stimuli sont générés en faisant varier l'entraxe caméra selon la règle de prise de vue optimale numéro 2. Pour la condition 5, la génération des stimuli a suivi la règle de prise de vue optimale numéro 1.

L'évaluation subjective de la QoE

Afin de vérifier les règles de prise de vue optimales proposées, un test subjectif a été réalisé pour étudier l'effet de la déformation des formes et de la zone de confort (5 scènes et 5 conditions de test) sur trois indicateurs de QoE comprenant l'expérience visuelle, le rendu de la profondeur et le confort visuel.

1. Stimuli : le matériel de test était composé de 5 scènes. Pour chaque scène, 5 stimuli différents correspondant à 5 conditions différentes ont été générés. Pour chaque session, 25 stimuli étaient utilisés.
2. Equipement : l'environnement de test était conforme avec la recommandation BT. 500 de l'UIT-R. Un écran S-3D entrelacé ligne de 46 pouces et d'une définition de 1920x1080 pixels a été utilisé pour la visualisation.
3. Observateurs : 28 observateurs non experts ont été recrutés. Un test de vision incluant l'acuité visuelle, l'acuité stéréoscopique, la fusion, la vision des couleurs etc. a été réalisée sur tous les testeurs pour déterminer leur performance visuelle. Les résultats montrent que tous les testeurs étaient capables de percevoir la profondeur binoculaire.
4. Procédure : la méthode SAMVIQ a été utilisée comme protocole de base. 3 indicateurs de QoE incluant le rendu de la profondeur, le confort visuel et l'expérience visuelle ont été évalués séparément dans des sessions de test différentes.

Analyse des résultats

Concernant le rendu de la profondeur, les résultats montrent que les sujets peuvent aisément distinguer les images stéréoscopiques des images 2D. La 2D a toujours été notée « médiocre ». Il n'est pas facile pour les sujets de distinguer le rendu de la profondeur pour les différents niveaux de profondeur 3D. Cependant, la condition 5 destinée à optimiser la forme pour la région d'intérêt montre un léger avantage comparée aux autres conditions.

Concernant le confort visuel, les résultats montent qu'il diminue avec l'augmentation de la valeur de DOF. Pour la plupart des scènes, les conditions 2 et 3 (DOF égale respectivement à 0,1 et 0,2), sont légèrement inférieures à la condition 2D. Cependant, le niveau de confort visuel est supérieur à 60 (bon à excellent). La condition 4 (DOF =

0,3) présente une diminution importante du confort visuel comparée à la condition 3 ce qui suggère que 0,2 dioptrie est le seuil de confort de visualisation.

Pour l'expérience visuelle, les stimuli 3D dont la plage de profondeur perçue est à l'intérieure de la zone de confort (0,2 dioptrie) sont tous notés au-dessus de la qualité « bonne ». Les images 2D sont jugées autour du niveau « assez bon », en dessous des images 3D confortables. La condition 4 qui se situe en dehors de la zone de confort est jugée « mauvaise » dans tous les cas.

Afin de comprendre la relation entre indicateurs subjectifs et pour prouver la priorité des règles proposées (règle de prise de vue numéro 3), nous avons regroupé les résultats subjectifs en 2 catégories : ceux dits « sans inconfort » (note MOS de confort visuel égale ou au-dessus de 60 « bon ») et ceux dits « avec problème d'inconfort » (note MOS en dessous de 60). L'analyse statistique révèle que dans le cas « sans inconfort », le rendu de la profondeur est le facteur dominant de l'expérience visuelle tandis que dans le cas « avec problème d'inconfort », le confort visuel est le facteur dominant pour l'expérience visuelle.

Principales conclusions et discussion

Plusieurs résultats et conclusions ont indiqué que les règles de prise de vue proposées et les priorités associées peuvent assurer une QoE visuelle optimisée (rendu de la profondeur, confort visuel et expérience visuelle) :

1) En ce qui concerne la règle 1, l'optimisation de la déformation des formes est démontré lors de l'évaluation du rendu de la profondeur et dans le cas « sans inconfort ». L'effet dominant du rendu de la profondeur sur l'expérience visuelle souligne également ce point.

2) En ce qui concerne la règle 2, les résultats montrent que la zone de confort de visualisation se situe, comme prévu, autour de 0.2 dioptries. Le confort visuel chute rapidement en-dessous de la qualité « bon » quand la scène restituée est en dehors de la zone de confort. Ainsi, garantir la profondeur perçue à l'intérieur de la zone de confort est nécessaire.

3) Les résultats confirment que la règle 2 est prioritaire par rapport à la règle 1. En présence d'un inconfort visuel, le confort visuel est le facteur dominant de l'expérience visuelle. De plus, on peut noter un lien étroit entre le rendu de la profondeur et le confort visuel. Ainsi, l'optimisation du confort visuel est prioritaire par rapport à l'optimisation de la distorsion de la forme.

R 6. L'impact de la variation de la profondeur binoculaire perçue sur la QoE en TV S-3D

Dans la section précédente, nous avons vérifié les règles de prise de vue stéréoscopiques proposées. Cependant, deux questions principales demeurent :

1) Seuls des contenus synthétiques ont été utilisés et les observateurs semblent avoir des difficultés pour juger le rendu de la profondeur de ce type de contenu. Ceci pourrait s'expliquer par le fait que les observateurs ne sont pas familiers avec les objets et le contenu des scènes synthétiques. Ainsi, il est important d'ajouter dans les stimuli des contenus naturels pour vérifier si les observateurs sont plus sensibles à la profondeur des contenus naturels.
2) Seulement trois indicateurs de QoE ont été utilisés. Ce nombre est peut-être insuffisant pour comprendre l'impact perceptuel des images stéréoscopiques.

Ainsi, il est important de concevoir un test subjectif avec un nombre d'indicateurs de QoE plus important, ainsi que des contenus naturels et synthétiques. Pour cela, le test subjectif a été construit en se focalisant sur l'exploration de la plus importante des valeurs ajoutées – les variations de la profondeur binoculaire – afin de déterminer comment celle-ci affecte la QoE des images stéréoscopiques. Pour chaque scène, les paramètres de prise de vue ont été choisis pour générer différents niveaux de profondeur binoculaire perçue. Six indicateurs de QoE comprenant la qualité d'image 2D, la quantité de profondeur, le confort visuel, le rendu de la profondeur, l'aspect naturel et l'expérience visuelle ont été utilisés. L'acceptabilité du confort visuel a également été mesurée sur une échelle binaire (acceptable ou non acceptable) afin de déterminer les critères d'acceptation des observateurs relativement au confort visuel en TV S-3D. En outre, l'étude du lien entre les différents indicateurs de QoE a permis de proposer un modèle perceptuel de QoE pour la TV S-3D.

R 6.1 Organisation de l'expérimentation

Un test subjectif a été organisé pour étudier l'effet de différentes scènes (2 scènes synthétiques et 3 scènes naturelles) et de différents niveaux de profondeur binoculaire perçue (3 niveaux de DOF produits en contrôlant les paramètres de prise de vue lors de la génération des images) sur six indicateurs de QoE incluant la qualité d'image 2D, la quantité de profondeur, le rendu de la profondeur, l'aspect naturel, l'expérience visuelle et le confort visuel. La description du test subjectif est présentée ci-dessous.

1. Stimuli : le matériel de test utilisé dans cette expérience était composé de 3 scènes naturelles et de 2 scènes synthétiques. Pour chaque scène, 4 stimuli différents correspondant à différentes plages de profondeur perçue de 0 / 0,1 / 0,2 / 0,3 dioptries ont été générés.
2. Equipement : l'environnement de test et l'écran étaient les mêmes qu'au chapitre 5. Cependant, la distance de visualisation a été changée à 4,5 fois la hauteur de l'écran afin d'augmenter la capacité de rendu de la profondeur et ainsi de fournir une meilleure sensation de profondeur aux observateurs.
3. Observateurs : 28 observateurs non experts ont été recrutés. Tous ont passé le même test de vision présenté au chapitre 5.
4. Procédure : la procédure de test était similaire à celle du chapitre 5 mais était composée de 6 sessions correspondant aux 6 indicateurs de QoE 3D. De plus, les sujets devraient reporter l'acceptabilité du confort visuel (échelle catégorielle à 2 niveaux : « acceptable », « non acceptable ») lors d'une session dédiée au confort visuel. La méthode SAMVIQ a été utilisée comme protocole de base.

R 6.2 Analyse des résultats

L'analyse statistique des résultats montre que la qualité d'image n'est pas affectée par la variation de la profondeur binoculaire. Le résultat relatif à la quantité de profondeur indique que les sujets peuvent facilement discriminer différentes plages de profondeur perçue. Avec l'augmentation de la profondeur perçue, le niveau de confort visuel diminue significativement. Le rendu de la profondeur, l'aspect naturel et l'expérience visuelle sont tous affectés de façon similaire par la variation de profondeur binoculaire. En augmentant légèrement la profondeur perçue, la 3D présente de meilleurs notes que la 2D (DOF = 0) qui est jugée « médiocre » pour le rendu de la profondeur et « assez bon » pour l'aspect naturel et l'expérience visuelle alors que la condition DOF=0,1 est jugée entre « bon » et « excellent » pour tous ces indicateurs. Cependant, quand la profondeur perçue est supérieure à une certaine valeur (DOF = 0,1 pour les scènes naturelles et DOF =0,2 pour les scènes synthétiques), l'écart de notes stagne, puis diminue et va jusqu'à s'inverser en faveur de la 2D. L'entretien auprès des observateurs, confirme que le confort de visualisation semble être le facteur prépondérant car l'avantage procuré par l'augmentation de la profondeur génère de l'inconfort visuel à partir d'une certaine valeur, faisant chuter alors, l'expérience visuelle, tout comme pour la déformation des formes.

En comparant les scènes naturelles et synthétiques en termes de quantité de profondeur et de qualité des images 2D, toutes les scènes se comportent de façon similaire. Pour les autres indicateurs de QoE, la note MOS des scènes naturelles chute plus vite que celle des scènes synthétiques.

Les résultats du test d'acceptabilité montrent qu'une acceptabilité de 80% correspond à une note de 60 pour le confort visuel, c'est-à-dire à la frontière entre « assez bon » et « bon ». Un taux d'acceptabilité de 50% correspond à une note de confort visuel de 50 (catégorie « assez bon »).

A partir de ces résultats, un modèle de QoE est proposé : des indicateurs de QoE 3D de haut niveau (rendu de la profondeur, expérience visuelle et aspect naturel) peuvent être définis comme la somme pondérée d'indicateurs bas niveaux (qualité image 2D, quantité de profondeur et confort visuel). Une analyse par régression linéaire simple a été réalisée sur les données de cette expérience. Les résultats de cet ajustement linéaire montrent la relation entre les indicateurs de haut niveau et ceux de bas niveau.

R 6.3 Principales conclusions et recommandation

Dans cette expérience, nous avons exploré comment la profondeur binoculaire affecte la qualité d'expérience offerte par la visualisation des images stéréoscopiques. Les résultats sont récapitulés ci-dessous :

- L'augmentation de la profondeur binoculaire augmente la quantité de profondeur perçue car les gens peuvent facilement juger différents niveaux de profondeur binoculaires. Cependant, en même temps, l'augmentation de profondeur binoculaire diminue le confort visuel.

- La qualité d'image 2D n'est pas affectée par la variation de la profondeur binoculaire.
- Il a été montré que les indicateurs de plus haut niveau de la QoE, le rendu de la profondeur, l'aspect naturel et l'expérience visuelle peuvent être prédits par une somme pondérée de la qualité d'image 2D, de la quantité de profondeur et du confort visuel quand seule la variation de la profondeur binoculaire est prise en compte. Le coefficient d'ajustement linéaire a prouvé que le confort visuel est le facteur dominant pour l'expérience visuelle (56.8%) et l'aspect naturel (54.1%).

De plus, à partir de ces résultats, des recommandations sur la production de contenus sont proposées :

- Pour les contenus synthétiques, une DOF de 0,2 dioptrie devrait être choisie pour maintenir le confort visuel.
- Pour les contenus naturels, une DOF de 0,1 dioptrie devrait être la cible.

R 7. Impact de la compression JPEG-2000 sur la QoE en TV S-3D

La compression utilise la redondance de l'information présente dans les images ou les séquences d'images pour réduire la quantité de données à stocker ou à transmettre. Comprendre l'impact de la technique de compression sur la QoE des images facilite le choix de la technique de compression et du débit optimal de transmission pour l'application choisie. Comme présenté au chapitre 1, les chaînes de diffusion TV S-3D actuelles tendent à réutiliser les techniques de compression conventionnelles pour comprimer et transmettre les signaux image stéréoscopiques.

L'impact de la technique de compression d'image sur la qualité d'image 2D a été bien étudié. Cependant, ces techniques pourraient ne pas être applicables directement à la S-3D TV. Il existe deux raisons principales à cela : d'abord, l'impact de nouveaux artéfacts (tels que les artéfacts binoculaires induits par la compression du contenu 3D) sur la qualité d'image 2D exige davantage de recherche ; en second lieu, la QoE de la TV S-3D est multidimensionnelle incluant non seulement la qualité d'image 2D mais également d'autres indicateurs de QoE tels que la quantité de profondeur et le confort visuel. L'impact de la technique de compression d'image sur la QoE de la TV S-3D reste encore à étudier.

Ainsi, dans ce chapitre, nous avons pour objectif d'étudier l'impact de la compression d'image sur la QoE d'images fixes stéréoscopiques. La compression JPEG-2000 a été employée comme système de compression. Les scènes de référence 3D comprenant deux scènes naturelles et deux scènes synthétiques ont été choisies parmi celles utilisées dans l'expérience du chapitre 6 afin d'éviter tout inconfort visuel issu de la production d'image. De plus, la vue gauche de chaque scène est employée pour représenter la version 2D. Les deux images du couple stéréoscopique sont codées en utilisant cinq niveaux différents de compression JPEG-2000. Un test subjectif avec cinq indicateurs de QoE (qualité d'image 2D, quantité de profondeur, confort visuel, rendu de la profondeur et expérience visuelle) a été construit pour évaluer l'impact de la compression JPEG-2000 sur la QoE des images 2D et 3D stéréoscopiques. L'analyse des résultats de cette étude permettra d'évaluer l'impact de la compression.

En outre, une hypothèse semblable au chapitre 6 - les indicateurs de niveau élevé de la QoE peuvent être estimés à partir des indicateurs de bas niveau de la QoE - est également évaluée dans cette étude.

R 7.1 Organisation de l'expérimentation

Le dispositif expérimental est semblable à celui de l'expérience présentée au chapitre 6. Notons que, d'une part, les observateurs ayant rencontré des difficultés pour évaluer l'aspect naturel de contenus synthétiques et d'autre part, afin de réduire la complexité du test subjectif, le concept « aspect naturel » n'a pas été évalué dans cette étude. Le test était destiné à étudier l'effet de la scène (4 scènes), de la dimension (2D, 3D) et du taux de compression JPEG-2000 (0, 500, 100, 175, 250) sur cinq indicateurs différents de QoE comprenant la qualité d'image, la quantité de profondeur, le confort visuel, le rendu de profondeur et l'expérience visuelle.

R 7.2 Analyse des résultats

La qualité d'image 2D décroit avec l'augmentation du taux de compression JPEG-2000 pour les images 2D et 3D. Cependant, les résultats montrent que la qualité d'image des séquences 3D diminue plus rapidement que celle des images 2D.

La quantité de profondeur des images 3D a été systématiquement mieux notée que celle des images 2D. La quantité de profondeur diminue légèrement avec l'augmentation du taux de compression à la fois pour les images 2D et les images 3D.

Pour les faibles taux de compression, le niveau de confort visuel pour la 2D et la 3D sont similaires. Cependant, la différence de niveau de confort visuel entre la 2D et la 3D croit avec l'augmentation du taux de compression. Les observateurs ont rapporté un niveau d'inconfort visuel plus élevé avec l'augmentation des distorsions dues à la compression.

Pour le rendu de la profondeur, les résultats sont similaires à ceux de la quantité de profondeur.

Concernant l'expérience visuelle, les images 3D sans aucune compression sont notées « excellent » alors de les images 2D sans aucune compression sont seulement jugées « bon ». Quand le taux de compression est inférieur à 100, l'expérience visuelle entre la 2D et la 3D est équivalente à une différence de 20 points MOS. Les images 3D apportent un avantage certain en expérience visuelle comparé à la 2D. Pour un taux de compression de 100, cet écart diminue d'environ 10 points sur l'échelle MOS. Pour des taux de compression supérieurs à 100, l'avantage des images 3D en expérience visuelle disparait car la note MOS 3D est évaluée équivalente à celle de la 2D.

De plus, une régression linéaire basée sur le modèle de QoE proposé (voir chapitre 5) a été réalisée pour mettre en évidence la relation entre les indicateurs de QoE de haut niveau et ceux de bas niveau.

R 7.3 Principales conclusions et recommandation

Les principaux résultats de cette étude peuvent être résumés de la façon suivante :

1) La compression JPEG-2000 fait baisser globalement la note de tous les indicateurs de QoE. En augmentant le taux de compression, la note MOS réduit de manière significative.
2) En comparant l'effet des défauts de compression sur les images 2D et 3D, l'expérience visuelle des images 3D est plus élevée que celle des images 2D quand le taux de compression est inférieur à 100. Cela peut s'expliquer par l'apport de la profondeur binoculaire. Cependant, cet avantage diminue avec l'augmentation du taux de compression. Ce qui peut s'expliquer par la présence d'artéfacts visuels supplémentaires en 3D induits par la distorsion de compression. Dans les taux plus élevés de compression, à la fois la qualité d'image et le confort visuel sont jugés avec des notes inférieures en 3D qu'en 2D.
3) L'avantage apporté par la quantité de profondeur entre la 3D et la 2D ne diminue pas significativement même pour les taux de compression élevés. Cela indique que la compression JPEG n'a pas détruit l'information de profondeur binoculaire.
4) Dans cette étude, les résultats issus de la régression linéaire ont montré que l'expérience visuelle peut être prédite à 34% par la qualité d'image, 28% par la quantité de profondeur et 38% par le confort visuel. De plus, le rendu de la profondeur peut être prédit à 87% par la quantité de profondeur et 13% par le confort visuel.

La principale recommandation de cette étude est que pour les taux de compression élevés ou les scénarios de services à bas débit, il n'y a aucun intérêt à fournir le service 3D. En effet, concernant l'expérience visuelle, les séquences 3D n'ont pas montré un avantage sur les séquences 2D. De plus, la 3D induit plus d'inconfort visuel et réduit plus la qualité d'image que dans le cas 2D.

R 8. Impact des formats de représentation d'image sur la QoE des écrans S-3D entrelacés ligne

Un des problèmes très importants de la chaîne de diffusion 3D est le choix du format de représentation S-3D. Deux vues pleine résolution représentent le choix idéal mais il génèrera probablement un trop fort volume de données, un débit très important et la plupart du temps inacceptable. Il est donc nécessaire de définir les nouveaux besoins qui conduiront à des normes adaptées et à de nouveaux équipements pour la compression et la diffusion. Par conséquent, différentes stratégies, telles que la demie résolution horizontale (format côte à côte), la demie résolution verticale (format dessus-dessous) sont employées dans l'industrie afin d'être compatibles avec les formats vidéo conventionnels de la TV HD. Cependant, leurs effets potentiels sur la qualité d'expérience en TV S-3D sont mal connus. De plus, la combinaison des formats de représentation, du mode de balayage vidéo (entrelacé ou progressif) et des techniques d'affichage des écrans TV S-3D (entrelacés ligne/colonne ou obturateur actif) peut affecter la qualité d'expérience finale.

Dans ce chapitre, nous avons pour objectif d'étudier l'influence des formats de représentation vidéo sur la qualité d'expérience perçue en considérant des écrans TV

3D entrelacés ligne. Dans cette étude, une partie des contenus sources sont en mode progressifs et les autres en mode entrelacés. Les vidéos de test ont été soigneusement choisies pour répondre à une règle de complexité vidéo 3D (texture, mouvement, profondeur). Différents formats de représentation S-3D avec différents niveaux de résolution horizontale, verticale ou diagonale (horizontale et verticale) ont été simulés.

Deux expérimentations ont été conçues : la première se concentre sur la comparaison directe de différents formats de représentation vidéo sans aucune compression et avec un écran 3D entrelacé ligne. L'évaluation de la performance de différents formats vidéo a été réalisée à l'aide d'un test subjectif basé sur la méthode SAMVIQ, utilisant deux indicateurs de QoE (l'expérience visuelle et le rendu de la profondeur) ; la deuxième expérience a comparé la QoE des formats vidéo côte à côte, dessus-dessous et 2D HD compressés à différents débits. Les résultats de ces études montrent l'impact des différents formats de représentation vidéo sur la QoE perçue en utilisant un écran entrelacé ligne.

Les différents formats de représentation 3D stéréoscopiques

La vidéo stéréoscopique devrait contenir les images destinées à l'œil droit et à l'œil gauche nécessitant ainsi le double de capacités pour le stockage des vidéos non compressées par rapport à la 2D. Même après la compression vidéo, le débit binaire de transmission peut encore être supérieur à la vidéo 2D conventionnelle. Aussi, et afin d'être compatible avec les formats TV HD actuels, la technique normalement utilisée consiste à réduire d'un facteur 2 les résolutions horizontale ou verticale. Dans ce chapitre, nous avons classé les formats 3D vidéos utilisés dans cette étude de la façon suivante :

1) Demie résolution horizontale pour chaque vue : correspond au format vidéo compatible « Side-by-Side » comme spécifié dans le document A154 de DVB. Par exemple, pour une image HD (progressive) de définition 1920x1080 pixels, la résolution de chaque vue est 960x1080 pixels.

2) Demie résolution verticale pour chaque vue : correspond au format vidéo compatible « Top-and-Bottom » comme spécifié dans le document A154 de DVB. Par exemple, pour une image HD (progressive) de définition 1920x1080 pixels, la résolution de chaque vue est 1920x540 pixels.

3) Résolution réduite à la fois en horizontal et en vertical pour chaque vue : ce format est semblable à une paire de vues pour un affichage autostéréoscopique. Par exemple, un affichage autostéréoscopique de 9 vues a normalement seulement un tiers des résolutions horizontale et verticale pour chaque vue ce qui signifie 640x360 pixels pour un écran natif HD de 1920x1080 pixels.

4) Pleine résolution pour chaque vue : ce format restitue 1920x1080 pixels pour chaque oeil dans le cas HD.

R 8.1 Expérimentation 1

Cette expérimentation a été réalisée pour étudier l'influence de différents formats vidéo non compressés (réduction de la résolution horizontale, réduction de la résolution verticale, réduction des résolutions horizontale et verticale et formats pleine résolution) sur deux indicateurs de QoE (l'expérience visuelle et le rendu de la profondeur en utilisant un écran S-3D entrelacé ligne.

Méthodologie

1. Stimuli : contenus progressifs au format 1080p25 et entrelacés au format 1080i25. Six scènes classées des faibles complexités aux fortes complexités en termes de texture, mouvement et profondeur ont été sélectionnées. Pour chaque scène, les stimuli suivants ont été générés :
   - Résolution horizontale : 0,66 (réduction de 1/3), 0,5 (Side-by-Side) et 0,375 (réduction de 5/8)
   - Résolution verticale : 0,66 (réduction de 1/3) et 0,5 (Top-and-Bottom)
   - Résolution mixte (réduction à la fois horizontale et verticale) : 0,44 (réduction de 1/3 en horizontal et vertical), 0,25 (1/2 en horizontal et vertical) et 0,11 (2/3 en horizontal et vertical)
2. Equipement et environnement de test : les mêmes que ceux utilisés dans le test subjectif du chapitre 5.
3. Observateurs : 28 observateurs ont été recrutés pour participer à ce test.
4. Procédure : le protocole était basé sur celui de la méthode SAMVIQ pour évaluer deux indicateurs de QoE dans des sessions de test différentes : l'expérience visuelle et le rendu de la profondeur.

Analyse des résultats

Concernant l'expérience visuelle, la note MOS diminue avec la réduction de la résolution image. Seules les références 3D cachée, la vidéo 2D et la réduction de résolution horizontale limitée à 0,5 (cas du format Side-by-Side) ont été évaluées comme « excellent ». De plus, l'expérience visuelle de la référence 3D cachée était considérée comme supérieure à la 2D. Pour un même taux de réduction de résolution, les stimuli ayant subi une réduction horizontale étaient jugés meilleurs que ceux ayant subi une réduction verticale.

Concernant le rendu de la profondeur, toutes les vidéos 3D étaient jugées meilleures que la 2D. De plus, le rendu de la profondeur décroit également avec la réduction de la résolution image. Les courbes relatives au rendu de la profondeur ont une forme semblable à celles de l'expérience visuelle. Pour un même taux de réduction de 0,5, la réduction horizontale permet un meilleur rendu de la profondeur que la réduction verticale.

Nous avons également comparé la note MOS des contenus entrelacés à celle des contenus progressifs. Pour les contenus progressifs, l'expérience visuelle des différents types de réduction de résolution reste proche de la qualité « excellent » quand les taux de réduction sont supérieurs à 0,5. Pour un même taux de réduction, les résolutions horizontales et verticales présentent un niveau d'expérience visuelle similaire. Pour les contenus entrelacés et pour la réduction horizontale, les courbes ont une forme semblable à celle des contenus progressifs. Cependant, dans le cas d'une réduction verticale, l'expérience visuelle diminue sérieusement avec la baisse de la résolution. Une explication possible est que, pour les contenus entrelacés, la résolution verticale par trame est seulement la moitié de la résolution horizontale. Ainsi, une réduction de résolution supplémentaire dans la direction verticale imposée par l'utilisation de l'écran entrelacé lignes affecte encore plus l'expérience visuelle en comparaison avec une réduction de résolution horizontale.

Discussion

La version 2D pleine résolution de chaque contenu, qui était intégrée au test sans que les testeurs le sachent, montre que l'expérience visuelle est inférieure à celle du même contenu 3D pleine résolution. Il est facile pour l'utilisateur de distinguer le contenu 2D et le contenu 3D puisque la 2D contient un rendu de la profondeur très pauvre.

Les résultats du test ont également démontré que la réduction de définition sur les yeux gauche et droit réduit la note MOS pour les deux indicateurs de QoE : l'expérience visuelle et le rendu de la profondeur. Il est donc important de préserver la définition de chaque image d'une paire stéréo pour éviter la dégradation de l'expérience utilisateur.

De plus, le format image associé à la technologie d'affichage 3D a un impact significatif sur la perception utilisateur. Pour l'écran entrelacé ligne utilisé dans cette étude, si le contenu est entrelacé, la réduction de résolution horizontale fournit une expérience visuelle meilleure que dans le cas de la réduction de résolution verticale. Ainsi, les résultats indiquent que le format « Side-by-Side » est meilleur que le format « Top-and-Bottom » lorsque le contenu entrelacé est joué sur un écran entrelacé ligne. Pour les contenus progressifs, il semble que les réductions de résolution horizontale et verticale offrent des niveaux semblables d'expérience visuelle malgré l'utilisation d'un écran entrelacé lignes. Ce résultat qui semble surprenant, pourrait être expliqué par le fait d'avoir utilisé des contenus progressifs synthétiques avec une faible complexité de texture, les rendant ainsi très peu sensibles à la réduction de définition verticale de l'écran. Des études supplémentaires avec plus de stimuli ayant différents niveaux de complexité de texture sont nécessaires pour consolider cette conclusion.

R 8.2 Expérimentation 2

Le test précédent a montré que la réduction de la résolution dégrade l'expérience visuelle et le rendu de la profondeur. Le format « Side-by-Side » est meilleur que le format « Top-and-Bottom » pour les contenus entrelacés restitués sur des écrans entrelacés ligne puisque avec la même définition effective, l'expérience visuelle est meilleure. La logique de cette deuxième expérimentation était de mieux comprendre les performances potentielles des formats compatibles HD « Side-by-Side », « Top-and-Bottom » et « 2D HD » présentés sur un écran entrelacé ligne avec différents débits de compression (4 débits vidéo), sur l'expérience visuelle.

Méthodologie

1. Stimuli : quatre scènes ont été sélectionnées de façon à avoir différentes complexités vidéo. Tous les contenus source étaient des contenus entrelacés au format 1080i25. De plus, trois types de formats vidéo ont été considérés dans ce test : "Side-by-Side", "Top-and-Bottom" et 2D HD. Pour chaque format, 4 débits de compression, 5Mbps, 8Mbps, 12Mbps et 16Mbps ont été générés en utilisant une solution hardware basée sur des codeurs et décodeurs H.264.
2. Equipement et environnement de test : identiques à ceux de l'expérimentation 1 de ce chapitre.
3. Observateurs : les mêmes observateurs que ceux de l'expérience 1 ont participé à l'expérience 2.
4. Procédure : similaire à celle de l'expérience 1 de ce chapitre. Elle était basée sur celle de la méthode SAMVIQ. Un seul indicateur de QoE a été évalué durant ce test : l'expérience visuelle.

Analyse des résultats

Comme dans l'expérience 1 de ce chapitre, même dans le cas des hauts débits de transmission à 16Mbps, le format "Top-and-Bottom" n'est jamais jugé au-dessus de « bon », tandis que les formats "Side-by-Side" et 2D HD sont tous les deux notés entre « bon » et « excellent ». Les conclusions confirment celles de l'expérience 1 : le format "Side-by-Side" est meilleur que le format "Top-and-Bottom" pour les contenus entrelacés présentés sur un écran S-3D entrelacé ligne. Les vidéos 3D non compressées étaient jugées « excellent » et meilleures que les formats "Side-by-Side" et 2D HD compressés à 16Mbps. Cependant, nous pouvons observer que l'écart en termes d'expérience visuelle entre le format "Side-by-Side" et la 2D HD augmente avec la réduction des débits de compression. Dans le cas des faibles débits de compression, comme 5 Mbps, la 2D HD reste autour de la qualité « bon » tandis que le "Side-by-Side" est jugé « médiocre ».

Discussion

A nouveau, cette expérience montre que le format « Side-by-Side » (demie définition horizontale) est plus approprié que le format « Top-and-Bottom » pour les contenus entrelacé affichés sur des écrans entrelacés ligne. Les résultats montrent que pour atteindre une expérience visuelle optimale, une sélection rigoureuse du format vidéo, adapté à la technique d'affichage 3D, est très importante. De plus, pour maintenir la même expérience visuelle que la 2D, la 3D peut exiger plus de débit. Dans ce test, nous pouvons observer que pour le format « Side-by-Side » un débit d'au moins 16Mbps est nécessaire pour atteindre un niveau d'expérience visuelle semblable à celui de la 2D.

R 8.3 Principales conclusions et recommandation

Les résultats de la première expérimentation montrent que le format "Side-by-Side" peut fournir une expérience visuelle meilleure que le format "Top-and-Bottom" pour les écrans entrelacés ligne, spécialement dans le cas de contenus entrelacés. Pour optimiser la qualité d'expérience en TV S-3D, les résultats démontrent que la sélection du format de représentation 3D devrait être effectué en tenant compte des technologies d'affichage 3D. Les résultats de la deuxième expérimentation montrent que pour maintenir le même niveau d'expérience visuelle, les contenus 3D diffusés dans un format image dit « compatible » requièrent plus de débit que les contenus 2D.

R 9. Impact de l'asymétrie de vues sur la QoE en TV S-3D

En TV S-3D, le problème d'asymétrie de vues peut provoquer un inconfort visuel sérieux et donc dégrader la qualité d'expérience. Dans (Kooi and Toet, 2004), les auteurs étudient la contribution relative des imperfections spatiales sur le confort visuel avec des paires d'image binoculaires qui peuvent causer un inconfort lors de la visualisation. A partir de tests subjectifs, les auteurs ont estimé et ont proposé un seuil pour différent type de manipulation binoculaire, par exemple la rotation, le grandissement, le décalage vertical ou encore la luminance. Les trois principales conclusions de leur étude sont:

1) Les facteurs qui déterminent le plus fortement le confort de visualisation stéréoscopique sont la disparité verticale, le crosstalk et le flou ;

2) La vision binoculaire de l'observateur a une influence très limitée sur le confort de visualisation binoculaire ;

3) L'Hyperstereopsis a un effet très faible sur l'inconfort visuel.

Cependant, dans leur recherche, il pourrait exister quelques problèmes potentiels susceptibles de modifier leurs conclusions :

1) la paire d'origine des images a été acquise directement à partir des caméras sans aucune post-production, c'est-à-dire qu'elles pourraient, dès l'origine, contenir un certain niveau d'asymétries entre vues ;

2) Le temps de présentation était seulement de 3 secondes par stimulus, c'est-à-dire que ce temps pourrait être trop court pour que l'observateur passe soigneusement en revue la totalité de l'image ;

3) L'environnement de visualisation, par exemple la distance de visualisation, la luminosité de l'arrière-plan, les deux projecteurs, n'étaient pas en accord avec la méthode normalisée ce qui peut affecter la reproductibilité et la fiabilité de l'expérience.

Dans cette étude, nous avons pour objectif d'étudier l'impact de l'asymétrie de vue sur la QoE en TV S-3D d'une manière plus critique : D'une part, les paires d'image d'origine sont soient des contenus synthétiques sans problème d'asymétrie ou des contenus naturels qui ont été corrigés en post-production de sorte que nous puissions exclure de potentielles asymétries entre vues des paires d'image originales ; d'autre part, l'expérience subjective a strictement suivie la méthode normalisée afin de garantir la reproductibilité et la fiabilité; L'expérience a permis d'estimer des seuils de visibilité et de gêne.

R 9.1 L'asymétrie de vues en TV 3D

Le problème d'asymétrie de vues peut être induit par différentes sources. Par exemple, en s'intéressant au procédé de création de contenu, la prise de vue, avec convergence physique des caméras, peut générer des disparités verticales et de la distorsion trapézoïdale, à cause de la structure géométrique. De même, le mauvais appariement de la position des caméras peut avoir pour conséquence un décalage vertical, une rotation, une différence de grandissement entre vues. Les différences de focale peuvent provoquer différents niveaux de flou et de grandissement. La désynchronisation des couleurs ou de la luminance sur différents capteurs peut induire des asymétries de couleur et de luminance. Pour le codage et la transmission, la stratégie de codage asymétrique peut produire plus d'artéfacts visuels sur une vue. Même le codage symétrique peut produire plus d'artéfacts visuels en profondeur. Côté visualisation, l'imperfection du filtre de l'écran ou des lunettes peut causer des asymétries en luminance, en couleur et des images fantômes (crosstalk), et un mauvais appariement de la position des projecteurs peut également produire des asymétries géométriques.

Dans cette étude, nous sélectionnons et résumons les principales asymétries entre vues généralement rencontrées dans trois groupes comprenant l'asymétrie de luminance, l'asymétrie de couleur et l'asymétrie géométrique afin de faciliter l'analyse.

L'asymétrie de luminance

L'asymétrie de luminance est l'asymétrie la plus commune en TV 3D. Elle peut être produite par la désynchronisation du niveau de luminance ou de la fonction gamma des caméras, les systèmes optiques additionnels tel que l'installation d'un miroir semi-transparent sur le système de caméras, l'imperfection du filtre de l'écran ou des lunettes. Pour être plus précis et plus pratique, les ingénieurs vidéo sont habitués à adapter le niveau de blanc et le niveau de noir entre les caméras afin d'éviter les asymétries de luminance entre les caméras. Ainsi, l'asymétrie de luminance concerne les asymétries de niveau de blanc et les asymétries de niveau de noir.

L'asymétrie de couleur

Côté système de production stéréoscopique ou système de visualisation, l'imperfection des filtres peut causer des asymétries de couleur, comme par exemple le filtre polarisé présent sur l'écran ou composant les lunettes. L'imperfection du triangle de couleur issu des caméras peut également induire une asymétrie de couleur sur chaque canal.

Et bien sûr, la technique de multiplexage des couleurs comme utilisé dans les techniques de visualisation stéréoscopique de type anaglyphe (couleurs rouge/cyan par exemple) cause un sérieux problème d'asymétrie de couleur. Dans cette étude, nous simulons l'asymétrie de couleur d'une manière semblable à l'asymétrie de luminance (couleur blanche) mais en la limitant à un seul canal de couleur.

L'asymétrie géométrique

L'asymétrie géométrique que nous étudions ici est liée à un alignement géométrique imparfait des vues stéréoscopiques due à un positionnement inapproprié des caméras ou des projecteurs aussi encore à une post-production inadéquate. L'asymétrie géométrique peut être induite par la configuration des caméras elle-même. Dans , les auteurs ont analysé la géométrie des caméras et des systèmes de visualisation stéréoscopiques afin de comprendre l'effet de la distorsion d'image dans les systèmes vidéo stéréoscopiques. Leur analyse a précisé que la prise de vue convergente peut engendrer une distorsion trapézoïdale et ainsi créer une disparité verticale sur le bord des images.

Dans cette étude, trois types d'asymétries géométriques sont simulées : le décalage vertical, la rotation de vue et le grandissement d'une vue.

R 9.2 Définition de l'expérimentation

Le premier objectif de cette étude est d'évaluer l'impact de l'asymétrie de vues sur la QoE en TV S-3D. Trois scènes, deux synthétiques et une naturelle ont été sélectionnées pour couvrir différentes complexités d'image (texture et profondeur perçue finale). Pour chaque type d'asymétrie entre vues, 4 niveaux de distorsion ont été générés. La gêne visuelle a été évaluée à l'aide d'une échelle de dégradation à 5 catégories et le confort de visualisation a été évalué grâce à une échelle continue à 5 niveaux de qualité, basée sur la recommandation UIT-R BT. 500. Ces évaluations ont permis de mesurer l'impact des asymétries entre vues sur la QoE en TV S-3D. La méthode de test était basée sur la méthode SAMVIQ.

1. Stimuli : trois paires d'images stéréoscopiques représentant différents niveaux de complexité d'image et ne présentant pas d'asymétries entre vues ont été sélectionnées comme contenus source. Toutes les images stéréoscopiques ont été sélectionnées et vérifiées avec attention à l'aide du logiciel « Pure » de la société StereoLabs afin de garantir l'absence d'asymétries entre vues dans les images originales. La sélection du niveau d'asymétrie des stimuli a été finalisée suite à un pré-test réalisé par trois experts Ceux-ci ont choisi quatre niveaux de distorsion parmi une grande quantité de stimuli originaux de manière à utiliser au maximum la dynamique de l'échelle de vote.
2. Equipement : l'environnement de test et l'écran étaient les mêmes que pour le test subjectif du chapitre 6.
3. Observateurs : 30 observateurs ont été recrutés pour participer au test subjectif.
4. Procédure : le test a été scindé en deux parties : la première partie incluant 8 tests correspondant à 8 types d'asymétries entre vues a été choisi pour évaluer les stimuli à partir d'une échelle de dégradation à 5 catégories sur la gêne visuelle; la

seconde partie incluant également 8 tests d'asymétries entre vues mais utilisant une échelle continue de qualité à 5 niveaux pour évaluer le confort visuel.

R 9.3 Résultats et recommandation

Dans cette étude, nous avons mené un test subjectif basé sur la méthode SAMVIQ et destiné à mesurer l'impact des asymétries entre vues sur la gêne visuelle et le confort visuel d'images fixes stéréoscopiques. Les résultats montrent que tous les types d'asymétrie de vues augmentent la gêne visuelle et réduisent le confort visuel si les asymétries sont importantes. Cependant, il est possible d'éviter la gêne visuelle et les problèmes d'inconfort visuel si les asymétries entre vues sont maintenues en dessous d'une certaine quantité.

Trois seuils, comprenant un seuil de visibilité, un seuil de gêne visuelle ainsi que d'un point de vue plus pragmatique un seuil d'acceptabilité (80 pour cent des observateurs acceptent le niveau de confort visuel) ont été estimés dans cette étude.. Il est intéressant de noter que pour trois types d'asymétries géométriques, nous avons mis en évidence que la disparité verticale maximum peut être employée comme indicateur commun puisque les observateurs sont plutôt plus sensibles à la disparité verticale maximale.

Tableau R- 2: seuils estimés pour les asymétries entre vues

	Seuil de visibilité	**Seuil de gêne visuelle**	**Seuil d'acceptabilité à 80%**
Asymétrie en luminance			
Niveau de noir	3 %	15%	11%
Niveau de blanc	11%	27%	20%
Asymétrie de couleur			
Niveau RVB	10%	20%	20%
Asymétrie géométrique			
Disparité verticale	0.39 %	0.9%	0.69%
Rotation	0,22 degré	0,63 degré	0,59 degré
Grandissement	0.55 %	1.7%	1.27%
Disparité verticale maximale	2,8 min	7 min	5,6 min

La plupart des seuils estimés présentés dans cette étude sont plus strictes que ceux proposés dans (Kooi and Toet, 2004). Ceci peut être expliqué de la manière suivante : En premier lieu, les stimuli ont été soigneusement choisis afin de couvrir uniformément l'ensemble de la dynamique de l'échelle d'évaluation, et, dans notre test, les niveaux de dégradations étaient plus nombreux pour chaque type d'asymétrie. Ainsi, il est possible de proposer un seuil plus précis, par exemple de 2,8 à 7 minutes d'arc pour le décalage vertical comparé au seuil de plus ou moins 34 minutes d'arc (1 PD) proposé dedans . En second lieu, dans notre test, le temps de présentation pour chaque stimulus était de 8 secondes avec la possibilité de rejouer à volonté la séquence comme cela est permis par la méthode SAMVIQ, alors que dans (Kooi and Toet, 2004), la durée de présentation pour passer en revue l'image était de 3 secondes seulement. Une plus longue présentation et plus de liberté pour passer en revue l'image permettent à l'observateur d'être plus critique sur les artéfacts visuels. Ainsi, nos résultats sont sans doute plus précis et restent fiables.

Les résultats obtenus dans cette étude et les seuils proposés permettent une prévision plus précise de la gêne visuelle et du problème d'inconfort visuel d'un système vidéo stéréoscopique. Cela aidera certainement à la conception et au choix des systèmes vidéo stéréoscopiques. Le seuil de visibilité peut être employé pour guider la conception d'un système vidéo stéréoscopique optimal. Le seuil d'acceptabilité à 80% devrait être employé pour les recommandations destinées aux services TV 3D car à ce jour, il n'existe aucun système optimal allant dans ce sens sur le marché. Ainsi, lors de la présence d'asymétries entre vues dans les systèmes stéréoscopiques, nos recommandations sont les suivantes:

- L'asymétrie de niveau de noir devrait être inférieure à 11% et celle de niveau blanc inférieure à 20%
- L'asymétrie de couleur, devrait être inférieure à 20%
- La disparité verticale, devrait être maintenue en dessous de 5,6 min d'arc.

Conclusion générale

Dans le chapitre 1, nous avons évalué les challenges à relever pour la QoE de la TV S-3D. Les challenges peuvent être classés en deux niveaux : le niveau perceptuel et le niveau technique. Au niveau perceptuel, nous avons expliqué que la TV S-3D devrait augmenter la perception de profondeur grâce à la présence de l'information binoculaire additionnelle qui représente un repère de profondeur sensible, tout en considérant des distances de visualisation typiques de la TV S-3D. Cependant, cette amélioration de la perception de profondeur pourrait être anéantie. En effet, les anomalies (par exemple, la dissociation de l'accommodation et de la convergence) lors de la visualisation en TV S-3D pourraient provoquer de l’inconfort visuel et de la fatigue visuelle. Au niveau technique, en passant en revue les différentes technologies présentes sur la chaîne de diffusion TV S-3D, nous constatons qu'il n'y a aucune technique transparente. De nombreux problèmes techniques existent qui pourraient avoir un impact potentiel sur la QoE finale.

Dans le chapitre 2, nous avons passé en revue l'état-de-art de l'évaluation subjective de la QoE pour la TV S-3D en se basant sur les recommandations de l'UIT et les études exploratoires présentées dans la littérature. Nous avons indiqué que les recommandations de l'UIT, comme par exemple la recommandation UIT-R BT.500, ne permettent pas d'évaluer correctement la QoE 3D, car les caractéristiques spécifiques apportées par la TV S-3D ne sont pas prises en compte. En ce qui concerne des études exploratoires, nous avons indiqué que les deux problèmes principaux sont le manque de définition précise des indicateurs de la QoE et le besoin de spécifications sur les environnements de visualisation. En se basant sur l’analyse ci-dessus, nous avons soulevé deux points essentiels pour développer une nouvelle méthode d'évaluation subjective de la QoE en TV S-3D : tout d'abord, la QoE de la TV S-3D est multidimensionnelle. Six indicateurs possibles (qualité d'image 2D, quantité de profondeur, confort visuel, rendu de la profondeur, aspect naturel et expérience visuelle) ont été définis pour évaluer l'effet à court terme de la QoE. Un indicateur particulier de QoE, la fatigue visuelle, a été défini pour évaluer l'effet à long terme de la QoE ; Ensuite, les nouveaux facteurs affectant l'évaluation de la QoE en TV S-3D ont été adressés et discutés, comme par exemple les spécifications de l'environnement de visualisation.

Dans le chapitre 3, nous nous sommes concentrés sur la spécification d'un des nouveaux facteurs essentiel pour l'évaluation subjective de la QoE en TV S-3D : la caractérisation des écrans TV S-3D en termes de rendu de la luminance et de performance du rendu de la profondeur. Pour le rendu de la luminance, nous avons proposé de mesurer et d'intégrer les nouvelles caractéristiques des écrans TV S-3D dans la définition des besoins relatifs à l'évaluation subjective de la QoE. Une étude de cas comparant la performance du rendu en luminance de différents écrans TV S-3D a été présentée. Les résultats montrent que le niveau de réduction de la luminance, la luminance finale perçue, la fonction gamma et le crosstalk changent selon l'écran. Ces éléments pourraient avoir unimpact potentiel sur la QoE finale et nécessitent des études supplémentaires pour validation. Dans cette thèse, la stratégie a été de choisir l'écran avec le meilleur rendu en luminance à partir de l'étude de cas. Pour le rendu de la profondeur, nous avons défini la capacité de rendu de la profondeur comme étant le nombre de plan en profondeur restitué dans la zone de confort de visualisation de l'écran TV S-3D, en combinant les paramètres physiques pour la géométrie de la profondeur et des paramètres perceptuels pour le confort visuel. Nous avons comparé la capacité de rendu en profondeur de différentes technologies d'écran. L'analyse des résultats a indiqué que la capacité de rendu de la profondeur des écrans dépend principalement de la distance de visualisation, de la taille des pixels, de la taille d'écran et de l'organisation des pixels de l'écran pour chacune des vues droite et gauche.

Dans le chapitre 4, nous avons conçu une expérience pour mesurer la fatigue visuelle lors de la visualisation d'un contenu vidéo d'une heure dans des conditions de visualisation optimales. Des contenus 2D et 3D ont été employés pour savoir s'il existe des différences relatives à la fatigue visuelle en 2D et en 3D. Trois méthodes, comprenant un questionnaire, un test de vision et la mesure d'un EEG ont été employées. Les résultats ont montré que concernant le questionnaire et le test de vision, il n'y a aucune différence significative de fatigue visuelle générée entre la 2D et la 3D. Ainsi, nous avons conclu que la visualisation 3D dans des conditions de visualisation optimales ne devrait pas entraîner de fatigue visuelle. Cependant, le résultat de la mesure d'EEG indique une différence significative pour la puissance des bandes bêta et gamma localisées dans les lobes frontal et postérieur du cerveau. Ceci pourrait refléter une différence dans le processus de perception de la profondeur par le cerveau entre la 2D et la 3D. Cependant, cela pourrait ne pas être nécessairement connexe à la fatigue visuelle.

Dans le chapitre 5, nous avons proposé des règles de prise de vue pour optimiser la QoE de la TV S-3D en considérant la déformation stéréoscopique et les contraintes permettant d'assurer une visualisation confortable. La déformation stéréoscopique des formes a été définie. Elle est basée sur un modèle géométrique reliant la perception de la profondeur binoculaire de l'espace des caméras à l'espace de visualisation. Plusieurs modèles de caméras et différentes configurations ont été analysés. Une zone de confort de visualisation a été définie à partir des seuils proposés dans la littérature. Notre proposition de règles de prise de vue optimales se compose de trois points : 1) adapter les paramètres de prise de vue ou les paramètres de la scène pour éviter la déformation stéréoscopique ; 2) adapter les paramètres des caméras ou les paramètres de scène pour garantir que la profondeur binoculaire perçue est maintenue dans la zone de confort de visualisation ; 3) garantir que le confort visuel est prioritaire par rapport à la déformation stéréoscopique. Une expérience subjective de QoE a été définie pour évaluer les règles de prise de vue proposées en utilisant trois indicateurs de QoE (l'expérience visuelle, le confort visuel et le rendu de la profondeur). Les résultats ont prouvé que les règles de prise de vue proposées et les priorités associées peuvent assurer une QoE optimisée.

Dans le chapitre 6, nous avons exploré l'impact de la variation de la profondeur perçue sur la QoE en TV S-3D. Les résultats ont prouvé que 1) augmenter la profondeur binoculaire perçue augmente la quantité de profondeur perçue mais diminue le confort visuel ; 2) la qualité d'image 2D n'est pas affectée par la variation de profondeur binoculaire perçue ; 3) des indicateurs de niveau élevé de la QoE comprenant le rendu de la profondeur, l'aspect naturel et l'expérience visuelle peuvent être prédits par une somme pondérée de la qualité d'image, de la quantité de profondeur et du confort visuel. D'ailleurs, des recommandations pour la production de contenus ont été proposées avec une DOF de 0,2 pour les contenus synthétiques et une DOF de 0,1 pour les contenus naturels.

Dans le chapitre 7, nous avons étudié l'effet de la compression JPEG-2000 sur la QoE en TV S-3D. Le résultat a montré que la compression génère une dégradation globalement pour tous les indicateurs de QoE. De plus, en comparant l'effet de la compression sur des images 2D et des images 3D, on constate que la 3D offre une meilleure expérience visuelle que pour la 2D, lorsque les contenus ne sont pas compressés ou pour des taux de compression faibles. Cependant, ce n'est plus le cas avec des forts taux de compression, qui génèrent une augmentation des artéfacts stéréoscopiques, ce qui pourrait faire baisser l'expérience visuelle. La conclusion de cette étude est qu'il n'y a pas d'intérêt à fournir un service 3D pour les bas débits puisque dans ce cas, la 3D fournit une plus mauvaise expérience visuelle que la 2D.

Dans le chapitre 8, nous avons exploré l'impact des formats de représentation d'image sur la QoE à partir de deux expériences réalisées avec un écran entrelacé ligne. Les résultats de la première expérience ont indiqué que le lien entre le format de représentation 3D et la technique d'affichage TV S-3D a un impact sur la QoE en TV S-3D. Le format « Side-by-Side » est mieux adapté à un affichage entrelacé ligne que le format « Top-and-Bottom ». La deuxième expérience a été conçue pour comparer les formats compatibles 2D (Side-by-Side et Top-and-Bottom) à différents débits de transmission. Les résultats confirment l'avantage du format « Side-by-Side » sur le format « Top-and-Bottom » pour un même débit de compression. De plus, les résultats ont également indiqué que pour maintenir la même expérience visuelle qu'en 2D HD, la 3D exige plus de débit. Dans cette étude, le débit recommandé pour diffuser de la 3D au format « Side-by-Side » et en utilisant la compression H.264 est d'au moins 16Mb/s.

Dans le chapitre 9, comme l'asymétrie entre vues est un problème particulier pour la 3D et qu'elle peut être générée à partir de différentes parties de la chaîne de diffusion TV S-3D, nous avons mesuré l'impact de l'asymétrie entre vues sur la QoE en TV S-3D. Trois types d'asymétries entre vues, comprenant la luminance, la couleur et la géométrie ont été simulées. Nos résultats ont confirmé que les asymétries entre vues induisent une gêne visuelle et causent de l'inconfort visuel. Afin d'optimiser la QoE en TV S-3D, nous recommandons que les asymétries entre vues d'un système vidéo stéréoscopique doivent être maîtrisées : 1) L'asymétrie de niveau de noir devrait être inférieure à 11% et de niveau de blanc à 20% ; 2) L'asymétrie de couleur devrait être inférieure à 20% ; 3) Pour la disparité verticale, elle devrait être maintenue à un niveau inférieur à 5,6 minutes d'arc.

Les contributions de cette thèse

La contribution de cette thèse couvre trois niveaux différents :

- Le premier niveau concerne les méthodologies d'évaluation de la QoE 3D. Nous avons proposé d'employer des indicateurs multidimensionnels pour mesurer la QoE en TV S-3D. Nous avons également mis en évidence de nouveaux facteurs affectant la QoE 3D lors des évaluations subjectives. Une partie de ces propositions a été soumise et acceptée dans le projet de recommandation P.3D-sam de l'UIT-T destiné à la normalisation des méthodes subjectives d'évaluation pour la qualité de la vidéo 3D.
- Le deuxième niveau est destiné à comprendre l'impact des problèmes perceptuels et techniques sur la QoE en TV S-3D. Plusieurs questions relatives à l'acquisition de contenus, aux formats de représentation 3D, à la compression d'image et au débit de transmission ont été adressées et étudiées.
- Le troisième niveau a pour objectif de fournir des recommandations pour optimiser la QoE en TV S-3D telles que proposées ci-dessous :
    - Les règles de prise de vue pour optimiser l'acquisition de contenus pour la TV S-3D. Ce travail a permis d'aboutir à un brevet déposé en France et à l'international.
    - Le budget de profondeur est fixé au maximum à 0,2 dioptries pour les contenus synthétiques et 0,1 dioptries pour les contenus naturels afin de garantir le confort visuel.
    - Le format « Side-by-Side » est le format compatible 2D approprié pour un affichage entrelacé ligne.
    - Un débit plus élevé que la 2D HD est nécessaire pour diffuser de la 3D en mode compatible 2D.

- Les seuils de visibilité et de confort visuel pour des asymétries de luminance, de couleur et de géométrie.

Appendix A: S-3D video encoding

Compared with 2D images, 3D images data does require more capacities for storage and transmission. Thus, the duty of compression is more important in order to reduce the amount of data. Furthermore, various types of 3D representation formats and their potential quality issues, e.g., the precision of the depth maps and the lack of occlusion layer, make the duty of compression even more challenging.

Conventional stereo video coding

Classical video coding methods such as MPEG 2 or H.264/MPEG-4 AVC (Richardson, 2003) can be directly used to compress conventional stereo video formats.

The simplest way is to multiplex views in one single 2D video frame, such as frame compatible formats like Side-by-Side and top-and-bottom. In this case, 3D videos are compressed in the same way as 2D videos.

Another method is called simulcast (shorthand for “simultaneous broadcast”) in which each view is encoded independent of the other. The advantages of simulcast are 1) low computation complexity since dependencies between views are not exploited; 2) backward compatibility since one of the views could be decoded for legacy 2D displays. The main drawback is the coding efficiency since redundancy between views is not considered.

MPEG 2 standard: the multi-view Profile (MVP) (Ohm, 1999) had been defined to facilitate the stereo two-views video coding. As shown in Figure A. 1, the left view is encoded as a key sequence, and the right view can be predicted from the left view. Both the temporal prediction and inter-view prediction are allowed. Thus, the coding efficiency can be improved while computation complexity may increase. A similar scheme is defined in H.264 standard: stereo high profile(Vetro et al., 2011). It achieves higher coding efficiency compared with the MEPG 2 – MVP because of many improvements in H.264, e.g., intra prediction, multiple reference frames, variable block size for the temporal prediction. Besides coding efficiency problem, L.Tseng et al. (L.Tseng and Anastassion, 1995) conduct an experiment applying a perceptual adaptive quantization approach to stereoscopic video coding. Their simulation results indicate the importance of perceptual stereo coding, with improvement in overall stereo quality and reduction in binocular artifacts.

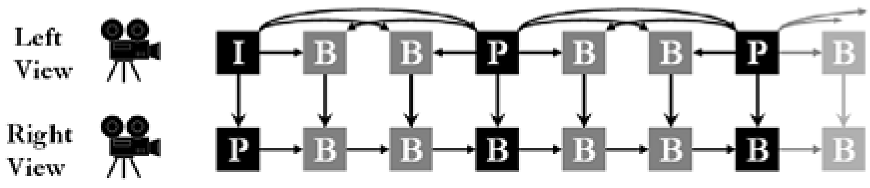


Figure A. 1 : Illustration of prediction in MPEG-2 Video MVP (Smolic et al., 2007)

2D-plus-depth coding

MPEG-C Part 3 (ISO/IEC 23002-3 Auxiliary Video Data Representation)(ISO, 2007) specified a standard for storage and compression of 2D-plus-depth data. The 2D image and the depth image are encoded independently, resulting in two separate coding streams. The depth image is compressed like conventional luminance signals using MPEG-2 or MPEG-4 video codecs with auxiliary container for depth information. The results from European project ATTEST (Fehn, 2003, Meesters et al., 2003b) claimed that the depth signal can be efficiently compressed by state-of-the-art video codecs (MPEG-2, MPEG-4, H.264/AVC). Because depth data are on average smoother and less structured than color data, it only required 10% to 20% of the bit rate of the 2D image to be encoded at good quality. However, the video-coding-induced distortion and depth-quantization-induced distortion (Liu et al., 2009) affect the quality of the view synthesis. Optimization of the coding algorithms by considering the human depth perception, e.g., depth quantization (Pastoor, 1992), is required.

Multiview video coding

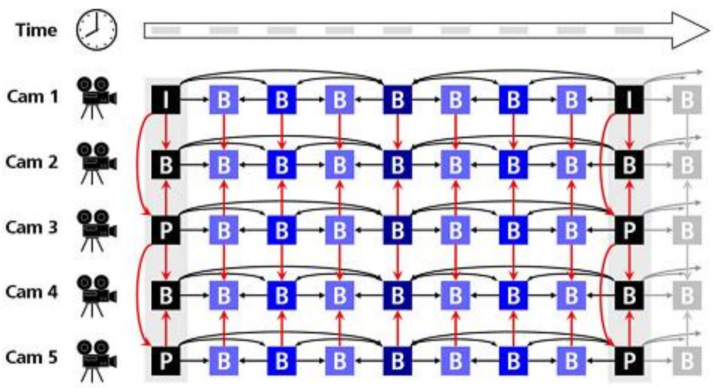


Figure A. 2 : Illustration of prediction in MVC (Smolic et al., 2007)

Multiview video coding (MVC) (Merkle et al., 2007b, Smolic et al., 2007, Merkle et al., 2007a) is an extension of the Advanced Video Coding (AVC) standard that provides efficient coding for MVV format. The main idea of this standard is to re-use MPEG-4 AVC encoding tools (hierarchical B images, temporal predictions and etc.) in order to reduce temporal and spatial redundancies contained in successive images(intra-view prediction) and adjacent video(inter-views prediction). As shown in

Figure A. 2, both temporal prediction and inter-view prediction are used to increase the coding efficiency. Compared with the MVP as shown in Figure A. 1 which only allowed prediction between 2-views and limited temporal prediction for the right view, MVC coding exploits all statistical dependencies with multi-view data set. For example, multi-references prediction is allowed in both temporal prediction of each view and inter-views prediction.

Merklet et al. in (Merkle et al., 2007b) showed that MVC outperformed the simulcast coding, with coding gain up to 3.2 dB and an average gain of 1.5 dB. They also stated two basic problems limiting the MVC coding efficiency: the first problem is large disparities between different views of MVV sequences and the second problem is inconsistencies of illumination and color across views.

Multi video plus depth coding

For multi-view video plus depth data (MVD), the current solution is to use MVC coding to encode the 2D image sequences and the depth images sequences independently (Merkle et al., 2007a). The relationship between the 2D color image and the depth is still under investigation. Thus, there are no coding standard which can take advantage of the dependencies between color texture image and the depth image to increase the MVD coding efficiency. Further research is required.

Coding for LDV and DES

European project 3D4YOU (Kerbiriou et al., 2010) investigated the coding method for LDV format. Their comparison between MVD and LDV format using MVC coding method (texture and depth sequences are encoded independently) showed that in normal camera baseline, MVD can provide better results than LDV with respect to the quality of rendered images (less artifacts). Advanced methods such as block alignment and temporal sub-sampling with data accumulation for occlusion layer were proposed to improve the coding efficiency and the quality of rendered images for LDV. For DES format, the coding method is still a widely open question.

Appendix B. Representation format conversion

Full resolution (progressive content) to line interleaved format

Figure B. 1 illustrates the process of converting full resolution (progressive content) left view and right view to line interleaved format for final representation. Firstly, a Bicubic filter (with low pass filter function) is used to resize each view to half vertical resolution. Then resized half resolution left view and right view are merged into final full resolution fame as left view in odd line and right view in even line. The reason of using Bicubic filtering (with low pass filter function) is to avoid aliasing.

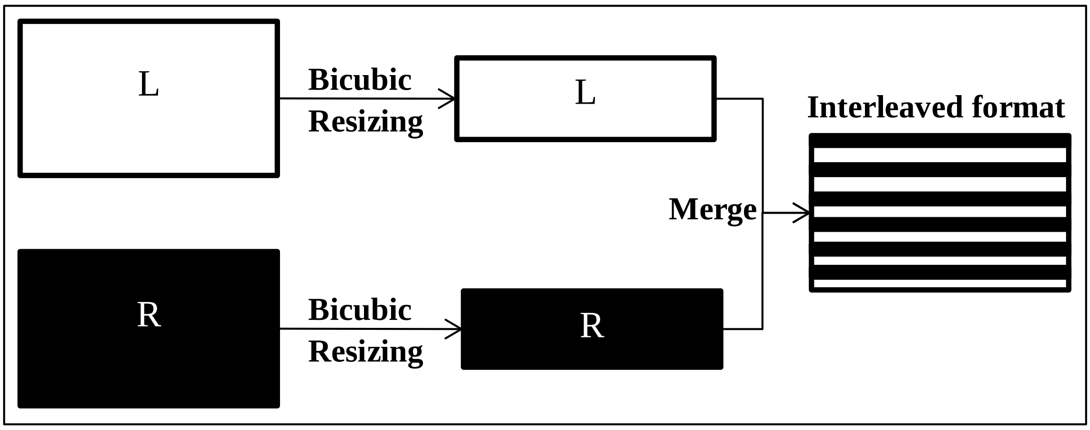


Figure B. 1 : Interleaved format conversion process for progressive content

Full resolution (interlaced content) to line interleaved format

Figure B. 2 illustrates the process of converting full resolution (interlaced format) left view and right view to line interleaved format for final representation. Each frame of interlaced content consists of two fields captured in different time (one after another) with half vertical resolution. Thus, two interlaced frames (from capture) from left and right view respectively will be converted to two interleaved frames representing different time stamp (for representation for display). Moreover, these two fields in each frame of one view have one pixels spatial shifting. Thus, firstly, a Bicubic upsampling is used to upsample two field of each view to be full resolution. This process is mainly to get rid of the one pixels spatial shifting between two fields in one interlaced frame. Then it is similar to the process as shown in Figure B. 1. Field 1 in left view and field 1 in right view will become frame 1 in the interleaved format. Field 2 in left view and field 2 in right view will become the frame after the frame 1 in the interleave format.

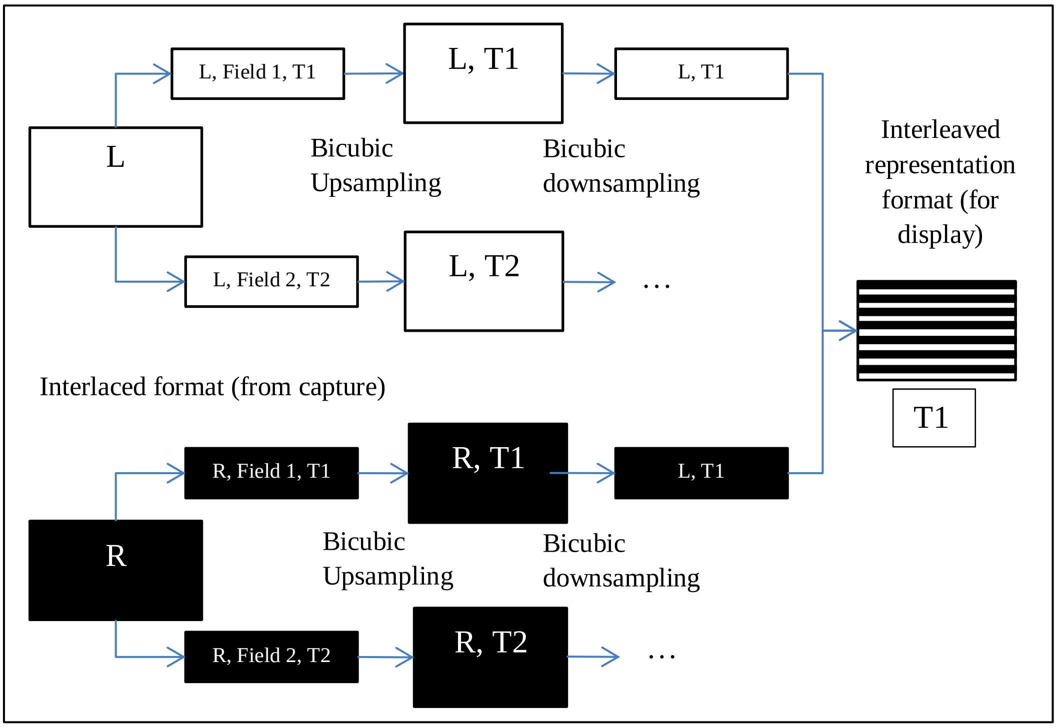


Figure B. 2 : Interleaved format conversion process for progressive content

Bibliography

n.d. *Mortal Journal: 3D Moive craze of the 1950's* [Online]. Available: http://www.mortaljourney.com/2010/11/1950-trends/3d-movies [Accessed July 2012].

ADAMS, M. D. n.d. *JasPer Project* [Online]. Available: http://www.ece.uvic.ca/~frodo/jasper/ [Accessed].

ALATAN, A. A., YEMEZ, Y., GUDUKBAY, U., ZABULIS, X., MULLER, K., ERDEM, C. E., WEIGEL, C. & SMOLIC, A. 2007. Scene Representation Technologies for 3DTV: A Survey. *Circuits and Systems for Video Technology, IEEE Transactions on,* 17**,** 1587-1605.

ANDREW, J. W. Year. Understanding Crosstalk in Stereoscopic Displays. *In:* 3DSA (Three-dimensional Systems and Applications) conference, 2010 Tokyo, Japan.

ARDITO, M., GUNETTI, M. & VISCA, M. 1996. Influence of display parameters on perceived HDTV quality. *Consumer Electronics, IEEE Transactions on,* 42**,** 145-155.

ATANAS BOEV, D. H., ATANAS GOTCHEV 2012. Classification of stereoscopic artefacts. *Project No.216503.* MOBILE3DTV.

BALTER, R., FOURNIER, J., GICQUEL, J.-C., KAPTEIN, R. & VINAYAGAMOORTHY, V. 2008. 3D4YOU WP4-Deliverable 4.1 Technical Requirements. *In:* WP4 (ed.). Philips,BBC,FT.

BANTON, T. & LEVI, D. M. 1991. Binocular summation in vernier acuity. *J. Opt. Soc. Am. A,* 8**,** 673-680.

BARKOWSKY, M., COUSSEAU, R. & CALLET, P. L. 2009. Influence of depth rendering on the quality of experience for an autostereoscopic display. *First International Workshop on Quality of Multimedia Experience.* San Diego, California, USA.

BATTIATOA, S., CURTIB, S., CASCIAC, M. L., TORTORAC, M. & SCORDATOC, E. Year. Depth-Map Generation by Image Classification. *In:* Three-Dimensional Image Capture and Applications VI, 2004 San Jose. 95.

BING, B. 2010. *3D and HD Broadband Video Networking,* Norwood, MA, USA, Artech House, Inc.

BLENDERFOUNDATION. 2008. *Big buck bunny* [Online]. Available: http://www.bigbuckbunny.org [Accessed].

BLENDERFOUNDATION. 2010. *Blender Ver 2.49B* [Online]. Available: www.blender.org [Accessed].

BLIN, J. L. Year. New quality evaluation method suited to multimedia context SAMVIQ. *In:* The Second International Workshop on Video Processing and Quality Metrics for Consumer Electronic, Jan 2006 Phoenix,Arizona.

BOEV, A., HOLLOSI, D., GOTCHEV, A. & EGIAZARIAN, K. Year. Classification and simulation of stereoscopic artifacts in mobile 3DTV content. *In:* ANDREW, J. W., NICOLAS, S. H. & JOHN, O. M., eds., 2009. SPIE, 72371F.

BOHER, P., LEROUX, T., BIGNON, T. & COLLOMB-PATTON, V. 2012. Full optical characterization of auto-stereoscopic 3D displays using local viewing angle and imaging measurements. 82880S-82880S.

BROADBENT, A. D. 2004. A critical review of the development of the CIE1931 RGB color-matching functions. *Color Research & Application,* 29**,** 267-272.

BRODMANN, K. 2006. *Brodmann's Localisation in the Cerebral Cortex: The Principles of Comparative Localisation in the Cerebral Cortex Based on Cytoarchitectonics,* New york, Springer.

BRULS, W. H. A., VAREKAMP, C., GUNNEWIEK, R. K., BARENBRUG, B. & BOURGE, A. Year. Enabling Introduction of Stereoscopic (3D) Video: Formats and Compression Standards. *In:* Image Processing, 2007. ICIP 2007. IEEE International Conference on, 2007. I - 89-I - 92.

BRUST, H., SMOLIC, A., MUELLER, K., TECH, G. & WIEGAND, T. Year. Mixed resolution coding of stereoscopic video for Mobile devices. *In:* 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, 2009, 4-6 May 2009 2009. 1-4.

BUSS, S. R. 2003. *3D Computer Graphics: A Mathematical Introduction with OpenGL,* New York, NY, USA, Cambridge University Press.

CHEN, J., ZHAO, J., WANG, X., HUANG, C., DONG, E., CHEN, B. & YUAN, Z. 2011. A Simple Semi-automatic Technique for 2D to 3D Video Conversion. *Artificial Intelligence and Computational Intelligence,* 7003**,** 336-343.

COUTANT, B. E. & WESTHEIMER, G. 1993. Population distribution of stereoscopic ability. *Ophthalmic and Physiological Optics,* 13**,** 3-7.

COZMAN, F. & KROTKOV, E. Year. Depth from scattering. *In:* Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on, 17-19 Jun 1997 1997. 801-806.

DELORME, A. & MAKEIG, S. 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods,* 9.

DINER, D. B. Year. A new definition of orthostereopsis for 3D television. *In:* Systems, Man, and Cybernetics, 1991. 'Decision Aiding for Complex Systems, Conference Proceedings., 1991 IEEE International Conference on, 1991. 1053-1058 vol.2.

DODGSON, N. A. Year. Variation and extrema of human interpupillary distance. *In:* Stereoscopic displays and virtual reality system XI, 2004 San Jose. 36-46.

DODGSON, N. A. 2005. Autostereoscopic 3D Displays. *Computer,* 38**,** 31-36.

DRASCIC, D. & MILGRAM, P. Year. Perceptual Issues in Augmented Reality. *In:* Stereoscopic Displays and Virtual Reality Systems III, 1996. 123-134.

DVB 2011. Digital Video Broadcasting (DVB): Frame Compatible Plano-Stereoscopic 3DTV (DVB-3DTV). *DVB Document A154.*

E.CUTTING, J. & M.VISHTON, P. 1995. Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about detph. *In:* W., E. & S., R. (eds.) *Handbook of perception and cognition.* San Diego: Academic Press.

EBU 2010. User requirements for Video Monitors in Television Production. *EBU – TECH 3320.* Geneva.

EBU, E. B. U.-. 2005. EBU Technical Recommendation R115 - 2005. *Furture High Definition Televsion Systems: The Need to Develop Television Production Equipment For A Progressive Scanned Image Format of 1920 Horizontal by 1080 Vertical Resolution at 50 And 60 Hz Frame Rates.* Geneva: EBU.

EMOTO, M., NOJIRI, Y. & OKANO, F. 2004. Changes in fusional vergence limit and its hysteresis after viewing stereoscopic TV. *Displays,* 25**,** 67-76.

FEHN, C. 2001. 3D TV Broadcasting. *In:* O.SCHREER, P.KAUFF & T.SIKORA (eds.) *3D Videocommunication.* John Wiley & Sons, Ltd.

FEHN, C. Year. A 3D-TV system based on video plus depth information. *In:* Signals, Systems and Computers, 2003. Conference Record of the Thirty-Seventh Asilomar Conference on, 2003. 1529-1533 Vol.2.

FLIERL, M. & GIROD, B. 2007. Multiview Video Compression. *Signal Processing Magazine, IEEE,* 24**,** 66-76.

FOURNIER, J. 1995a. *ETUDE DE LA QUALITE DES IMAGES STEREOSCOPICQUES EN TELEVISION.* Doctor, UNIVERSITE DE RENNES I.

FOURNIER, J. 1995b. *Evaluation de la qualité visuelle perçue des images stéréoscopiques en télévision.* PHD, Unvierstiy of Rennes 1.

FOURNIER, J. & THIERRY, A. 1994. Human Factors requirements for a stereoscopic television service: Admissible contrast differnces between the two channels of a stereoscopic camera. *Symposium on Electronic Imaging Science & Technology* San Jose.

FROEHLY, L., MARTIN, S. N., LASSER, T., DEPEURSINGE, C. & LANG, F. 2003. Multiplexed 3D imaging using wavelength encoded spectral interferometry: a proof of principle. *Optics Communications,* 222**,** 127-136.

GAUTIER, J., BOSC, E. & MORIN, L. 2010. Representation and coding of 3D video data. *Project PERSEE - SCHEMAS PERCEPTUELS ET CODAGE VIDEO 2D ET 3D.*

GOLDMANN, L., DE SIMONE, F. & EBRAHIMI, T. Year. A comprehensive database and subjective evaluation methodology for quality of experience in stereoscopic video. *In:* BASKURT, A. M., ed., 2010a San Jose, California, USA. SPIE, 75260S-11.

GOLDMANN, L., JONG-SEOK, L. & EBRAHIMI, T. Year. Temporal synchronization in stereoscopic video: Influence on quality of experience and automatic asynchrony detection. *In:* Image Processing (ICIP), 2010 17th IEEE International Conference on, 26-29 Sept. 2010 2010b. 3241-3244.

GOLDMANN, L., SIMONE, F. D. & EBRAHIMI, T. Year. Impact of acquisition distortions on the quality of stereoscopic images. *In:* Fifth International Workshop on Video Processing and Quality Metrics for Consumer Electronics - VPQM 2010, 2010c Scottsdale, Arizona, U.S.A.

GOODWIN, C. 1995. Seeing in Depth. *Social Studies of Science,* 25**,** 237-274.

HDMI 2009. High-Definition Multimedia Interface. *Specification Version 1.4 Extraction of 3D Signaling Portion.*

HIRUMA, N., HASHIMOTO, K. & TAKEDA, T. Year. Evaluation of stereoscopic display task environment by measurement of accommodation responses. *In:* Engineering in Medicine and Biology Society, 1996. Bridging Disciplines for Biomedicine. Proceedings of the 18th Annual International Conference of the IEEE, 1996. 672-673 vol.2.

HODGES, L. F. & DAVIS, E. T. 1993. Geometric Considerations for Stereoscopic Virtual Environments. *GVU Technical Report;GIT-GVU-93-02.*

HOKODA, S. C. & CIUFFREDA, K. J. 1983. Theoretical and Clinical Importance of Proximal Vergence and Accomodation. *In:* SCHOR, C. M. & CIUFFREDA, K. J. (eds.) *Vergence Eye Movements: Basic and Clinical aspects.* Boston: Butterworth.

HOLLIMAN, N. 2004a. *3D Display Systems,* IOP Press.

HOLLIMAN, N. S. Year. Mapping perceived depth to regions of interest in stereoscopic images. *In:* WOODS, A. J., MERRITT, J. O., BENTON, S. A. & BOLAS, M. T., eds., 2004b San Jose, CA, USA. SPIE, 117-128.

HYUNG-CHUL, O. L., JUNHO, S., KEETAEK, K. & SEUNGHYUN, L. Year. Method of Measuring Subjective 3-D Visual Fatigue: A Five-Factor Model. *In*, 2008. Optical Society of America, DWA5.

IDDAN, G. J. & YAHAV, G. Year. 3D IMAGING IN THE STUDIO(AND ELSEWHERE...). *In:* SPUE Videometrics and Optical Methods for 3D shape Measurements, 2001 San Jose, CA, USA. 48-55.

IJSSELSTEIJN, W. A., DE RIDDER, H. & VLIEGEN, J. 2000. Subjective evaluation of stereoscopic images: effects of camera parameters and display duration. *Circuits and Systems for Video Technology, IEEE Transactions on,* 10**,** 225-233.

IJSSELSTEIJN, W. A., SEUNTIËNS, P. J. H. & MEESTERS, L. M. J. 2002. State-of-the-art in human factors and quality issues of stereoscopic broadcast television. Eindhoven University of Technology Department Technology Management.

IKEDA, M. & NAKASHIMA, Y. 1980. Wavelength difference limit for binocular color fusion. *Vision Research,* 20**,** 693-697.

INOUE, T. & OHZU, H. 1997. Accommodative responses to stereoscopic three-dimensional display. *Appl. Opt.,* 36**,** 4509-4515.

ION-PAUL, B. & HANNA, B. 1990. Influence des distorsions géométriques dans les images stéréoscopiques sur le confort visuel de l'observateur. *La Télévision en Relief.* CCETT Rennes.

ISO 2007. Information technology - MPEG video technologies Part 3: Representation of auxiliary video and supplemental information. Geneva, Switzerland: ISO.

ITU 1998. SUBJECTIVE ASSESSMENT METHODS FOR IMAGE QUALITY IN HIGH-DEFINITION TELEVISION. *Recommendation ITU-R BT.710-4.* International Telecommunication Union.

ITU 1999. Subjective video quality assessment methods for multimedia applications. *Recommendation P.910.* International Telecommunication Union.

ITU 2000. Subjective Assessment of Stereoscope Television Pictures. *RECOMMENDATION ITU-R BT.1438.*

ITU 2002. Methodology for the subjective assessment of the quality of television pictures. *Recommendation BT 500-11.* International Telecommunication Union.

ITU 2007a. Methodology for the subjective assessment of video quality in multimedia applications. *Question ITU R 102/6.*

ITU 2007b. Specifications and alignment procedures for setting of brightness and contrast of displays. *Rec. ITU-R BT.814-2.*

JANTET, V., MORIN, L. & GUILLEMOT, C. Year. Incremental-LDI for Multi-View Coding. *In:* Proceedings of the 3DTV-Conference 2009, The True Vison, Capture, Transmission and Display of 3D Video, Postdam, Germany, 2009. 1-4.

JENS OGNIEWSKI, I. R. 2011. Autostereoscopy and Motion Parallax for Mobile Computer Games Using Commercially Available Hardware. *International Journal of Computer Information Systems and Industrial Management Applications,* 3**,** 8.

JIMÉNEZ, J. R., OLIVARES, J. L., PÉREZ-OCÓN, F. & DEL BARCO, L. J. 2000. Associated Phoria in Relation to Stereopsis with Random-Dot Stereograms. *Optometry & Vision Science,* 77**,** 47-50.

JING LI, M. B., JUNLE WANG, PATRICK LE CALLET 2011. Study on Visual Discomfort induced by Stimulus Movement at Fixed Depth on Stereoscopic

Displays using Shutter. *17th International Conference on Digital Signal Processing 2011.* Greece

JOLLY, S., ZUBRZYCKI, J., GRAU, O., VINAYAGAMOORTHY, V., KOCH, R., BARTCZAK, B., FOURNIER, J., GICQUEL, J.-C., TANGER, R., BARENBRUG, B., MURDOCH, M. & KLUGER, J. 2009. 3D Content Requirements & Initial Acquisition Work. *WP1 – Deliverable 1.1.2.* 3D4YOU project

JONES, G., LEE, D., HOLLIMAN, N. & EZRA, D. 2001. Controlling Perceived Depth in Stereoscopic Images. *Proc. SPIE Stereoscopic Displays and Virtual Reality Systems VIII,* 4297**,** 42-53.

JUNG, Y. J., BAIK, A., KIM, J. & PARK, D. Year. A novel 2D-to-3D conversion technique based on relative height depth cue. *In:* WOODS, A. J., HOLLIMAN, N. S. & MERRITT, J. O., eds. Stereoscopic Displays and Applications XX, 2009 San Jose. 72371U.

KALVA, H., CHRISTODOULOU, L. & FURHT, B. Year. Evaluation of 3DTV Service using Asymmetric View Coding Based on MPEG-2. *In:* 3DTV Conference, 2007, 2007. 1-4.

KAUFF, P., MÜLLER, M., ZILLY, F., SMOLIC, A., VAREKAMP, C., BARENBURG, B., KLEIN-GUNNEWIEK, R., BRULS, F., BRIAND, G., KERBIRIOU, P., BOISSON, G., KLUGER, J., MALCHOW, W.-N., KOCH, R., BARTCZAK, B. & GRAU, O. 2008. 3D4YOU Deliverable D2.1.2: Requirement on post-production and formats conversion. *In:* WP2 (ed.). Philips,CAU,KUK,BBC,TMMRDF,HHI.

KERBIRIOU, P., COLLEU, T., MUELLER, K., GUNNEWIEK, R. K., VLEUTEN, R. V. D., RELIER, S. & GRAU, O. 2010. Comparative study and recommendations. 3D4YOU.

KOOI, F. L. & TOET, A. 2004. Visual comfort of binocular and 3D displays. *Displays,* 25**,** 99-108.

KOZAMERNIK, F., SUNNA, P., WYCHKENS, E. & PETTERSEN, D. I. 2005. Subjective quality of Internet video codecs - Phase 2 evaluation using SAMIVQ. *EBU technical Review.*

KUBOTA, A., SMOLIC, A., MAGNOR, M., TANIMOTO, M., CHEN, T. & ZHANG, C. 2007. Multiview Imaging and 3DTV. *Signal Processing Magazine, IEEE,* 24**,** 10-21.

KUZE, J. & UKAI, K. 2008. Subjective evaluation of visual fatigue caused by motion images. *Displays,* 29**,** 159-166.

L.TSENG, B. & ANASTASSION, D. 1995. Perceptual Adaptive Quantization of Stereoscopic Video coding using MPEG-2 temporal scalability structure. *International Workshop on Stereoscopic and Three Dimensional Image* Santorini.

LAMBOOIJ, M., FORTUIN, M., IJSSELSTEIJN, W. A. & HEYNDERICKX, I. 2009a. Measuring Visual Discomfort associated with 3D Displays. *Proceedings of SPIE – Stereoscopic Displays and Applications XX,* 7237.

LAMBOOIJ, M., IJSSELSTEIJN, W., BOUWHUIS, D. G. & HEYNDERICKX, I. 2011. Evaluation of Stereoscopic Images: Beyond 2D Quality. *Broadcasting, IEEE Transactions on,* 57**,** 432-444.

LAMBOOIJ, M., IJSSELSTEIJN, W., FORTUIN, M. & HEYNDERICKX, I. 2009b. Visual Discomfort and Visual Fatigue of Stereoscopic Displays: A Review. *Journal of Imaging Science and Technology,* 53**,** 030201-14.

LAMBOOIJ, M., IJSSELSTEIJN, W. & HEYNDERICKX, I. 2007. Stereoscopic displays and visual comfort: a review. SPIE Newsroom: Einhoven University of Technology Einhoven Univeristy of Technology Philips Research.

LI, H. C. O., JUNHO, S., KEETAEK, K. & SEUNGHYUN, L. Year. Measurement of 3D Visual Fatigue Using Event-Related Potential (ERP): 3D Oddball Paradigm. *In:* 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, 2008, 2008. 213-216.

LIU, Y., HUANG, Q., MA, S., ZHAO, D. & GAO, W. 2009. Joint video/depth rate allocation for 3D video coding based on view synthesis distortion model. *Signal Processing: Image Communication,* 24**,** 666-681.

LOH, A. M. & HARTLEY, R. Year. Shape from non-homogeneous, nonstationary, anisotropic, perspective texture. *In:* the British Machine Vision Conference 2005. 69--78.

MACCHIARELLA, P. 2010. Trends in 3DTV. Parks associates.

MALMIVUO, J. & PLONSEY, R. 1995. *Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields,* New York Oxford, OXFORD UNIVERSITY PRESS.

MANGIONE, C. M. 2000. The National Eye Institute 25-ItemVisual Function Questionnaire (VFQ-25).

MANGIONE, C. M., LEE, P. P., GUTIERREZ, P. R., SPRITZER, K., BERRY, S. & HAYS, R. 2001. Development of the 25-Item National Eye Institute Visual Function Questionnaire. *Archives of Ophtalmology,* 119**,** 1050-1058.

MEAD, B. 2010. The return of 35mm 3D: Two competing processes add a new dimension to stereoscopic race. Available: http://www.filmjournal.com/filmjournal/content_display/news-and-features/features/technology/e3if3d775aa9c43698eab6fb2500f376296.

MEESTERS, L., IJSSELSTEIJN, W. A. & SEUNTIENS, P. Year. A survey of perceptual quality issues in three-dimensional television systems. *In:* Stereoscopic Displays and Virtual Reality Systems X, 2003a Santa Clara, CA, USA. SPIE, 313-326.

MEESTERS, L. M. J., IJSSELSTEIJN, W. A. & SEUNTIENS, P. J. H. 2004. A survey of perceptual evaluations and requirements of three-dimensional TV. *Circuits and Systems for Video Technology, IEEE Transactions on,* 14**,** 381-391.

MEESTERS, L. M. J., QUANTE, B., SEUNTIËNS, P. J. H. & IJSSELSTEIJN, W. A. 2003b. Report on Formative Evaluations of Software Prototype Chain and Display Systems.

MENDIBURU, B. 2009. 3D Movie Making: Stereoscopic Digital Cinema From Scrip to Screen.

MERKLE, P., SMOLIC, A., MULLER, K. & WIEGAND, T. Year. Efficient Compression of Multi-View Depth Data Based on MVC. *In:* 3DTV Conference, 2007, 2007a. 1-4.

MERKLE, P., SMOLIC, A., MULLER, K. & WIEGAND, T. 2007b. Efficient Prediction Structures for Multiview Video Coding. *Circuits and Systems for Video Technology, IEEE Transactions on,* 17**,** 1461-1473.

MIKŠÍCEK, F. 2006. Causes of Visual Fatigue and Its Improvements in Stereoscopy. *Technical Report No. DCSE/TR-2006-04.* University of West Bohemia in Pilsen.

MITSUHASHI, T. & YUYAMA, I. Year. Recent Advances in HDTV - Present State and Future Development of HDTV. *In:* Image Management and

Communication (IMAC) in Patient Care: New Technologies for Better Patient Care, 1991. The Second International Conference on, 1991. 212-217.

MOOTE, S. & LENNON, C. 2010. 3-D TV: Are you ready? *Brodcast Engineering.*

NAGATA, S. 1996. The binocular fusion of human vision on stereoscopic displays - field of view and enviroment effects. *Ergonomics,* 39**,** 9.

NTT 2011. 3D quality assessemtn methods. NTT.

OHM, J.-R. Year. Stereo/multiview video encoding using the MPEG family of standards. *In:* Stereoscopic Displays and Virtual Reality Systems VI, 1999 San Jose, CA, USA. 242-253.

OWENS, D. A. & LEIBOWITZ, H. W. 1983. Perceptual and motor consequences of tonic vergence. *In:* SCHOR, C. M. & CIUFFREDA, K. J. (eds.) *Vergence Eye Movements: Basic and Clinical Aspects.* Boston: Butterworths.

PALMER, S. 1999. *Vision Science: Photons to Phenomenology*, The MIT Press.

PAPPAS, T. N. & SAFRANEK, R. J. 2000. Perceptual Criteria for Image Quality Evaluation. *Handbook of Image and Video Processing.* Academic Press.

PARK, C. S. & PARK, H. W. 2001. A robust stereo disparity estimation using adaptive window search and dynamic programming search. *Pattern Recognition*, 34**,** 2573-2576.

PASTOOR, S. Year. Human factors of 3DTV: an overview of current research at Heinrich-Hertz-Institut Berlin. *In:* Stereoscopic Television, IEE Colloquium on, 1992. 11/1-11/4.

PATTERSON, R. 2007. Human factors of 3-D displays. *Journal of the Society for Information Display,* 15**,** 10.

PHILLIPS, S. & STARK, L. 1977. Blur: A sufficient accommodative stimulus. *Documenta Ophthalmologica,* 43**,** 65-89.

REDERT, A. 2005. *Creating a Depth Map*. the Netherlands patent application.

RICHARDS, W. 1970. Stereopsis and stereoblindness. *Experimental Brain Research,* 10**,** 380-388.

RICHARDSON, I. E. G. 2003. *H.264 and MPEG-4 Video Compression Video Coding for Next-generation Multimedia*, WILEY.

ROGOWITZ , B. E., L AND T. N. PAPPAS Year. The rehabilitation of gamma. *In:* Human Vision and Electronic Imaging III, 1998 San Jose, CA, USA. 232-249

RUO, Z., PING-SING, T., CRYER, J. E. & SHAH, M. 1999. Shape-from-shading: a survey. *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* 21**,** 690-706.

SANEI, S. & CHAMBERS, J. A. 2007. *EEG Signal Processing,* Wiley-Interscience.

SAXENA, A., NG, A. & CHUNG, S. 2005. Learning Depth from Single Monocular Images. *NIPS*, 18.

SAYGILI, G., GURLER, C. G. & TEKALP, A. M. 2011. Evaluation of Asymmetric Stereo Video Coding and Rate Scaling for Adaptive 3D Video Streaming. *Broadcasting, IEEE Transactions on,* 57**,** 593-601.

SCHOR, C. M. & TYLER, C. W. 1981. Spatio-temporal properties of Panum's fusional area. *Vision Research,* 21**,** 683-692.

SCHOR, C. M. & WOOD, I. 1983. Disparity range for local stereopsis as a function of luminance spatial frequency. *Vision Research,* 23**,** 1649-1654.

SEUNTIËNS, P. 2006. *Visual experience of 3D TV.* doctor doctoral thesis, Eindhoven University of Technology.

SEUNTIËNS, P., MEESTERS, L. & IJSSELSTEIJN, W. 2006. Perceived quality of compressed stereoscopic images: Effects of symmetric and asymmetric JPEG coding and camera separation. *ACM Trans. Appl. Percept.,* 3**,** 95-109.

SEUNTIENS, P. J., HEYNDERICKX, I. E., IJSSELSTEIJN, W. A., VAN DEN AVOORT, P. M. J., BERENTSEN, J., DALM, I. J., LAMBOOIJ, M. T. & OOSTING, W. Year. Viewing experience and naturalness of 3D images. *In:* Three-Dimensional TV, Video, and Display IV, 2005 Boston, MA, USA. SPIE, 601605-7.

SEUNTIËNS, P. J. H., MEESTERS, L. M. J. & IJSSELSTEIJN, W. A. 2005. Perceptual attributes of crosstalk in 3D images. *Displays,* 26**,** 177-183.

SHADE, J., GORTLER, S., HE, L.-W. & SZELISKI, R. 1998. Layered depth images. *Proceedings of the 25th annual conference on Computer graphics and interactive techniques.* ACM.

SHIBATA, T., KIM, J., HOFFMAN, D. M. & BANKS, M. S. Year. Visual discomfort with stereo displays: effects of viewing distance and direction of vergence-accommodation conflict. *In:* Stereoscopic Displays & Applications XXII, 2011a San francisco, US.

SHIBATA, T., KIM, J., HOFFMAN, D. M. & BANKS, M. S. 2011b. The zone of comfort: Predicting visual discomfort with stereo displays. *Journal of Vision,* 11(8):11**,** 29.

SMOLIC, A., MUELLER, K., MERKLE, P., KAUFF, P. & WIEGAND, T. Year. An overview of available and emerging 3D video formats and depth enhanced stereo as efficient generic solution. *In:* Picture Coding Symposium, 2009. PCS 2009, 6-8 May 2009 2009. 1-4.

SMOLIC, A., MUELLER, K., STEFANOSKI, N., OSTERMANN, J., GOTCHEV, A., AKAR, G. B., TRIANTAFYLLIDIS, G. & KOZ, A. 2007. Coding Algorithms for 3DTV-A Survey. *Circuits and Systems for Video Technology, IEEE Transactions on,* 17**,** 1606-1621.

SOLUTIONS, P. D. 2008. 3D Interface Specifications, White Paper. Eindhoven, The Netherlands: Philips Electronics Nederland B. V. .

SPIRO, L. n.d. *A Brief History of Stereographs and Stereoscopes* [Online]. Available: http://cnx.org/content/m13784/latest/ [Accessed July 2012].

STARK, L., KENYON, R. V., KRISHNAN, V. V. & CIUFFREDAI, K. J. 1980. Disparity vergence: a proposed name for a dominant component of binocular vergence eye movements. *American journal of optometry & physiological optics,* 57**,** 3.

STROHMEIER, D., JUMISKO-PYYKKÖ, S. & KUNZE, K. 2010. NEW, LIVELY, AND EXCITING OR JUST ARTIFICIAL, STRAINING, AND DISTRACTING : A Sensory profiling approach to understand mobile 3D audiovisual quality. *Fifth International Workshop on Video Processing and Quality Metrics for Consumer Electronics - VPQM 2010.* Scottsdale, Arizona, U.S.A.

SURYAKUMAR, R. 2005. *Study of the dynamic interactions between vergence and accommodation.* Doctor of Philosophy, University of Waterloo.

TANIMOTO, M. 2006. Overview of free viewpoint television. *Signal Processing: Image Communication,* 21**,** 454-461.

UKAI, K. & HOWARTH, P. A. 2008. Visual fatigue caused by viewing stereoscopic motion images: Background, theories, and observations. *Displays,* 29**,** 106-116.

UKAI, K., TORII, M., WATANABE, M. & YAMADA, S. 2009. Accommodative responses to conflicted accommodation-convergence stimuli of stereoscopic retinal scan display.

VAN EE, R. & RICHARDS, W. 2002. A planar and a volumetric test for stereoanomaly. *Perception,* 31**,** 51-64.

VETRO, A., WIEGAND, T. & SULLIVAN, G. J. 2011. Overview of the Stereo and Multiview Video Coding Extensions of the H.264/MPEG-4 AVC Standard. *Proceedings of the IEEE,* 99**,** 626-642.

WATT, A. 1999. *3d Computer Graphics,* Boston, MA, USA, Addison-Wesley Longman Publishing Co., Inc.

WEI, Q. 2005. Converting 2D to 3D: A Survey. TU Delft.

WHEATSTONE, C. 1838. Contribution to the Physiology of Vision: Part the first, On some remarkable, and hitherto unobserved, Phenomena of Binocular Vision. *Philosophical Transactions 128*.

WHEATSTONE, C. 1850. Contributions to the Physiology of Vision.Part II. On Some Remarkable, and Hitherto Unobserved, Phaenomena of Binocular Vision,(Continued). [Abstract]. *Abstracts of the Papers Communicated to the Royal Society of London,* 6**,** 138-141.

WOODS, A. J. Year. Optimal Usage of LCD Projectors for Polarised Stereoscopic Projection. *In:* Stereoscopic Displays and Virtual Reality Systems VIII, 2001 San Jose, CA, USA Andrew J. Woods, Mark T. Bolas, John O. Merritt, Stephen A. Benton

WOODS, A. J., DOCHERTY, T. & KOCH, R. Year. Image distortions in stereoscopic video systems. *In:* Stereoscopic Displays and Applications IV, 1993 San Jose, CA, USA. SPIE, 36-48.

WÖPKING, M. 1992. Viewing comfort with stereoscopic pictures: An experimental study on the subjective effects of dispairty magnitude and depth of focus. *Journal of the Society for Information Display,* 3**,** 3.

YAMAGISHI, K., KARAM, L., OKAMOTO, J. & HAYASHI, T. Year. Subjective characteristics for stereoscopic high definition video. *In:* Quality of Multimedia Experience (QoMEX), 2011 Third International Workshop on, 7-9 Sept. 2011 2011. 37-42.

YAMANOUE, H. Year. The Differences Between Toed-in Camera Configurations and Parallel Camera Configurations in Shooting Stereoscopic Images. *In:* Multimedia and Expo, 2006 IEEE International Conference on, 2006. 1701-1704.

YAMANOUE, H., NAGAYAMA, M., BITOU, M. & TANADA, J. Year. Orthostereoscopic conditions for 3D HDTV. *In:* BOLAS, M. T., FISHER, S. S. & MERRITT, J. O., eds., 1998 San Jose, CA, USA. SPIE, 111-120.

YAMANOUE, H., OKUI, M. & OKANO, F. 2006. Geometrical analysis of puppet-theater and cardboard effects in stereoscopic HDTV images. *Circuits and Systems for Video Technology, IEEE Transactions on,* 16**,** 744-752.

YANO, S., EMOTO, M. & MITSUHASHI, T. 2004. Two factors in visual fatigue caused by stereoscopic HDTV images. *Displays,* 25**,** 141-150.

YANO, S., IDE, S., MITSUHASHI, T. & THWAITES, H. 2002. A study of visual fatigue and visual comfort for 3D HDTV/HDTV images. *Displays,* 23**,** 191-201.

YEH, Y.-Y. & SILVERSTEIN, L. D. 1990. Limits of fusion and depth judgment in stereoscopic color displays. *Hum. Factors,* 32**,** 45-60.

YONG-SHENG CHEN, J.-H. C., CHU-SONG CHEN,YI-PING HUNG, CHAN-HUNG SU, CHIOU-SHANN FUH 2001. Video-based eye tracking for autostereoscopic displays. *Optical Engineering* 40**,** 2726.

ZIOU, D. & DESCHENES, F. 2001. Depth from Defocus Estimation in Spatial Domain. *Computer Vision and Image Understanding,* 81**,** 143-165.

ZONE, R. 2007. *Stereoscopic Cinema & the Origins of 3-D Film, 1838-1952*, The university press of kentucky.